2024 IEEE 74th Electronic Components and Technology Conference (ECTC) | 979-8-3503-7598-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ECTC51529.2024.00072

Moving Towards Microchannel-based Chip Cooling

Paul Semenza
SEMI
Milpitas, USA
psemenza@semi.org

Dave Thomas
KLA
Newport, UK
dave.thomas@kla.com

Garrett Oakes
EV Group
Tempe, USA
G.Oakes@EVGroup.com

Dave Kirsch
EV Group
Tempe, USA
d.kirsch@evgroup.com

Yin Hang
Meta
Menlo Park, USA
yinhang@meta.com

Kuo-Chung Yee
TSMC
Hsinchu, Taiwan
kcyee@tsmc.com

Michael Cumbie
HP Inc.
Corvallis, USA
michael.cumbie@hp.com

Paul Benning
HP Inc.
Corvallis, USA
paul.benning@hp.com

Madhusudan Iyengar
Google
Mountain View, USA
miyengar@google.com

Lihong Cao
ASE
Fremont, USA
lihong.cao@aseus.com

William Chen
ASE
Fremont, USA
william.chen@aseus.com

Gity Samadi
SEMI
Milpitas, USA
gsamadi@semi.org

***Abstract*—Results of a review of the cooling needs for high performance computing (HPC) processors using advanced packaging indicates benefits of implementing direct cooling through silicon microchannels. Progress in fabrication of silicon microchannels, both in microcooler/cold plate and in direct cooling of the die is assessed. Challenges in transferring silicon microchannels into HPC packaging are summarized, and steps towards commercialization are presented.**

Keywords—high-performance computing, advanced packaging, microchannels, direct cooling

I. Introduction

The Si Microchannels study was convened by SEMI to identify materials, processes, and technologies that improve thermal management, performance, and yields for high performance computing (HPC) processors. The study sought to identify reasons for the lack of commercialization of silicon microchannels. A goal of the study was to issue a "call to action" to the industry to move silicon microchannels into production. The rest of this paper provides the motivation for adopting microchannel-based cooling, comments on relevant research and demonstration projects, a summary of key challenges for putting such approaches into production, and recommendations for actions that can be taken by industry.

II. HPC Chip Packages: Increasing Complexity and Thermal Loads

Advanced packaging has become a critical aspect of high-performance computing (HPC) solutions. The cost of system-on-chip (SoC) designs, which integrate functions such as central processing unit (CPU), graphic processing unit (GPU), and memory, into a single chip have dramatically increased as chip size and complexity have increased. In HPC applications, heterogenous integration has been adopted, in which SoCs are redesigned into smaller chiplets and integrated using advanced packaging. Heterogenous integration also enables mixing chiplets from different fabs and process nodes, which can lower cost and diversify supply.

The package types of particular interest for HPC applications include 2.1D, which utilizes thin-film or silicon bridges to connect chips; 2.3D, which uses silicon or organic interposer layers; 2.5D, in which passive interposer layers with through-silicon vias (TSVs) are used; 3D, in which chips are stacked on an active interposer with TSVs; and combinations of these [1].

These advanced packaging approaches worsen the thermal problem by increasing system power density, leading to higher heat fluxes and thermal crosstalk between adjacent or stacked dice, which are in close physical, and thermal, proximity. These challenges cannot be addressed by traditional air-cooled heat sinks. Air cooling is highly constrained by the physical limitation of the system design; most systems are limited by efficiency and cost considerations to less than 500W, as air cooling above this level incurs significant increases in fan power, airflow budget, heatsink size and cost, and noise, tradeoffs determined by the data center capabilities and restrictions.

Cold plate designs (single or dual-phase) have been adopted for increased thermal loads. Immersion cooling has also been used to cool multiple, densely-spaced boards; this can result in energy savings, but requires re-design of the system. Immersion cooling can handle ranges of 0.5-1.0 W/mm^2; immersion uses low-boiling point fluid, and is generally passive.

Remote cooling via attached microcoolers with engineered thermal paths transferring heat away from the chip to the ambient air or fluid enables improve cooling. The combination of high-power chips, high heat flux hotspots, and the high heat density resulting from chip stacking and 3-D integration, have created cooling challenges beyond the capability of Gen-2 approaches. New thermal management approaches capable of dealing with high module-level heat densities (W/cm^3) and high chip heat fluxes (W/cm^2) require inward migration of thermal management techniques, enabling heat removal by flowing liquids as close as possible to the on-chip heat source.

Microfluidic cooling can provide far lower junction-to-ambient thermal resistances than traditional cooling techniques in small volumetric footprint. Approaches for implementing this

2377-5726/24/$31.00 ©2024 IEEE
D

include cold plates with microchannels and integrating heat sinks into the backside of the dice (each high-power tier in the 3D stack may need its own backside heat sink), and interposer level cooling without modification to the mounted ICs.

While air-cooled heat sinks yield 100–130 W per chip, embedded cooling can reach ten times that dissipation, i.e., 1 to 1.5 kW, through the use of high-conductivity substrates and vias; dielectric liquids; directing liquids through manifolded structures; moving from single phase to evaporative cooling; manifolding the distribution of the evaporating fluid and directing the fluid across shorter flow paths, and to hotspots; or on-chip thermoelectric devices [2].

The development and implementation of such thermal management technologies was the focus of the DARPA Intrachip Enhanced Cooling Fundamentals (ICECool) thermal packaging program launched in 2013. This program aimed to develop and demonstrate "embedded cooling" techniques capable of removing kW/cm^2 chip heat fluxes and kW/cm^3 chip stack heat densities, while suppressing the temperature rise of multi-kW/cm^2 submillimeter hotspots. ICECool-funded projects explored embedded thermal management, bringing microfluidic cooling inside the substrate, chip or package and by including thermal management in the earliest stages of electronics design.

III. Efforts to Date in Liquid Chip Cooling

A. Microchannel Cold Plates

Hung, et al. reported using a CoWoS package with a 1.5X reticle size TSV interposer on a 60 mm x 60 mm organic substrate with a logic processor and four mimic HBM2 integrated through fine-pitch flip-chip bonding for thermal performance measurement with liquid cooling. Modeling of results indicates thermal design power up to 2kW is possible [3].

Cold-plate cooling has been adopted in production HPC packages, and has been evolving to reduce thermal resistance. In Fig. 1, the left-hand configuration involves a lead around the chip package, separated from the die by a thermal interface material layer (TIM1). The cold plate is built in a case, and is separated from the package lead by another layer (TIM2). The design of the channels inside the cold plate is critical to maximize heat removal while minimizing pumping power. In general, system/end users design the cold plate, and integrate it with the chip package/lead produced by the chip packaging supplier. Thermal resistance in this design is dominated by TIM2, which can be ~1mm thick, whereas TIM1 is 100 µm or thinner; TR can be lowered by making TIM2 thinner or better conducting. This is a mature process, and enables an arms-length relationship between chip packaging and system assembly.

The middle configuration removes TIM2 by integrating the cold plate at assembly. The packaging supplier creates a rim around the package, and the system house integrates the cold plate. This configuration has been adopted. The right-hand configuration is a focus of development currently. In this case, the cooling liquid contacts the back side of the silicon, further lowering the TR. The lack of a fully-contained cold plate can lead to fluid leaking into the package, resulting in shorts.

This configuration requires integration with the package, which implies that it will need to be done by foundry/OSAT, so co-design is required. The integration of the package into the system means that mechanical, electrical, and fluid interfaces need to be designed.

Several advanced designs for cold plates embedding microchannels have been reported, including using deep reactive-ion etching (DRIE) in silicon to create high aspect-ratio microchannels in a manifold design [4, 5]; fabricating fluidic microbumps and vias, electrical microbumps, and micropin-fin heat sink on a silicon die [6]; direct liquid cooling using Si-based jet impingement micro-coolers with jet slot, drain slot, and multiple pin-fin structures [7]; and a stacked silicon microcooler design with five layers [8]. Additive manufacturing of cold plates has also been explored, for example stereolithographic 3D printing of a clear polymer to fabricate a microfluidic jet impingement cooling package [9].

B. Microchannels on Back Side of Device

In the first reported experiment applying microchannels to cool ICs, Tuckerman and Pease demonstrated dissipation of 790 W/cm^2 using a simple orientation-dependent etch process on silicon wafers. The etch process was able to create channels of 50 x 300 microns, and the authors speculated that higher aspect ratio channels would enable heat dissipation exceeding 1kW/cm^2 [10]. Phillips later reported experiments with InP microchannel heat sinks with heat dissipation rates as high as 1,056 W/cm^2 [11]. Numerous studies have been conducted in recent decades, but the technology has not been used in production. DARPA funded research through the Near Junction Thermal Transport (NJTT) and ICECool programs, and the shift to 2.5D, 3D, and other advanced packaging approaches, have led chip makers, packaging houses, and other semiconductor supply chain participants to re-engage with this approach.

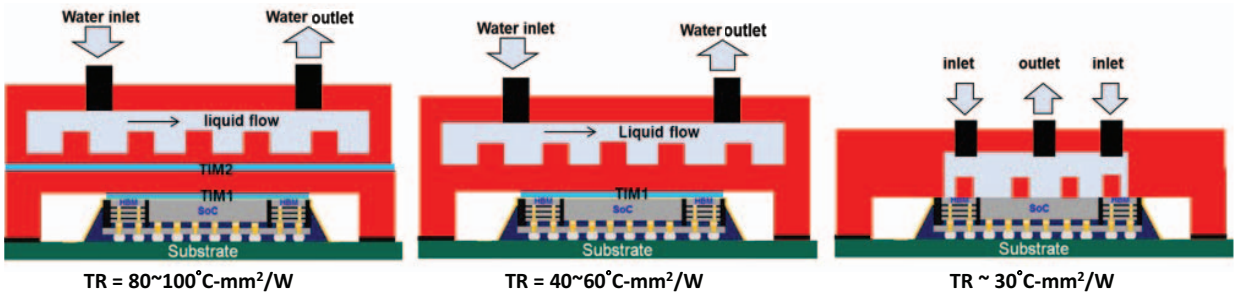

Fig. 1. 2.5D liquid cooling trends (source: TSMC)

Schultz, et al. reported modifying microprocessor modules through the creation of an embedded channel design followed by the development of a module fabrication and assembly process. Core power densities (over a 3.6 mm x 4.8 mm area representing a microprocessor core) of over 350 W/cm^2 were shown to be feasible along with local (200 μm x 200 μm) hot-spot power levels of more than 2 kW/cm^2 [12]. Sarvey, et al. integrated a micropin-fin heat sink into the back of a flip chip bonded silicon die [13]. van Erp, et al. demonstrated direct cooling of GaN-on-Si power ICs using the Si substrate as a microchannel heat sink. A device level heat flux of 417 W/cm^2 was cooled with 60 mW of pumping power [14]. Table 1 provides selected results from the research described above.

TABLE I. RESULTS OF CHIP COOLING RESEARCH

Design	*Heat flux dissipation (W/cm^2)*	*Reference*
Hierarchical manifold microchannel cold plate	1020	Drummond, et al. (2018)
3D manifold cold plate	250	Jung, et al. (2019)
Stacked silicon microcoolers	136-182	Bergendahl, et al. (2020)
Liquid jet impingement 3D polymer heatsink	189[a]	Wei, et al. (2020)
Direct lidded cooling with micropin array	133[b]	Hung, et al. (2021)
Microchannels on back side of die	790	Tuckerman, Pease (1981)
Microchannels on back side of die (InP)	1056	Phillips (1988)
Radial micro-channels with micro-pin fins on back side of die	350	Schultz, et al. (2017)
Manifold multichannel in the back side of GaN-on-Si	417	Van Erp, et al. (2020)

[a.] reported extrapolation to 1000W; chip size assumed to be 23x23mm

[b.] reported extrapolation to 2000W; assumption of 26x32mm SoC and 6 HBM stacks of 11x10mm for a total chip area of 15cm^2

IV. SI MICROCHANNEL DESIGN AND MANUFACTURABILITY

A. Microchannel Architecture

The cooling capability of microchannel heat sink is determined by the coolant thermal conductivity, the mean temperature difference between the substrate and the coolant, the area of the microchannels divided by the diameter of the microchannels, and the coolant flow rate. Thus, increasing the thermal conductivity of the coolant, lowering the temperature of the coolant, increasing the area of the microchannels and decreasing the diameter of the channels, and increasing the flow rate all lead to higher amounts of heat removal. Using two-phase cooling, in which the coolant is allowed to boil in the channels, further increases the amount of heat that can be removed, or allows the same amount of heat removal at lower flow rates, reducing the pumping power required. Maintaining fluid flow in two-phase operation is more complex, as pressure drop along the length of the channels leads to large pumping power requirements at small channel diameters and high flow rates. Manifold microchannels introduce the working fluid at multiple locations along the length of the channels, resulting in multiple flow paths of decreased effective flow length that can help with pressure losses. Two-phase evaporative cooling can also improve surface temperature uniformity and heat dissipation efficiency relative to single-phase cooling, and can enable reductions in size, weight, and overall power consumption. However, two-phase systems pose many challenges, such as the instability of rapid bubble growth, parallel channel flow, and upstream compressible volume [15]. These instabilities are hard to predict and come on abruptly, often leading to sudden, chaotic performance fluctuations and ultimately device failure. To make two-phase microfluidic systems commercially viable, it will be necessary to develop improved models.

HPC chips in production have interposers that exceed the size of typical wafer photo reticle (25 mm x 33 mm), with plans to continue to extend this trend with 2.5D packaging up to 100 mm x 100 mm and in some cases larger. Microchannel designs will need to mirror this approach, and will need to scale up. It will be important to work with industry to standardize a roadmap around next generation SoC package form factors and size in order to establish development guidelines for silicon microchannel cooling. Given the manufacturing challenges and limits of 2.5D heterogeneous chip packages (yield, warp/bow), both 2.5D and 3D approaches will be applied in future chip packages to continue to optimize the chip performance.

The three approaches to chip cooling envisioned are: at the top, within, or at the bottom of the stack (Fig. 2). Top of die stack cooling is easiest to imagine, as the top of current packages are accessible for cooling through conventional heat sinks. Within the die stack is the most useful, as it brings cooling inside the SoC processors where heat generation is the highest, but is also the hardest from an integration perspective, as it is necessary to etch channels directly into the prime die. Cooling at the bottom of the stack is more challenging, as I/O and power delivery must share the same space as fluidic cooling, trading off one performance constraint for another. For 3D ICs, fluid routing between two die in a stack is a special form of bottom of stack cooling, but will require improved interposer and joining technologies. No EDA tools exist today to help circuit architects understand these trade-offs to design approaches to delivering cooling, such as microfluidic interposers that are capable of both routing electrical interconnect through the substrate and channeling fluid through the substrate.

Cold plates are typically made by machining channels into metal with CNC or scything processes; the state-of-the-art approach is split-channel [16]. These channels are typically hundreds of microns in width, several millimeters in height, and tens of millimeters in length. This has the benefit of very low pressure drops in the microchannel achieved with minimal manifolding complexity. However, due to the relatively long length of these channels, heat transferred into the coolant saturates leading to poor heat transfer efficiency.

Delivering fluid through parallel channels needs to consider both the thermal cross talk between “hot” and “cold” channels (routing in and out), in addition to the effects of microchannel loses and thermal uniformity down the column.

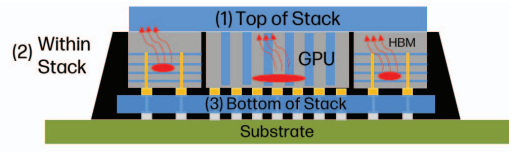

Fig. 2. Parallel microchannel consideration

Heat sink designs must balance flow rates and uniformity, temperature uniformity, and heat transfer efficiency while maintaining reasonable pressure drops. Pin-fin designs have been studied to increase heat transfer by maximizing surface area while maintaining a low pressure drop, but have been shown to lack thermal uniformity [17]. Another novel approach to address thermal uniformity or hot-spots that might arise was uses a hydrogel pin that swells or shrinks based on temperature, to dynamically allocate flow through constrained channels where heat flux is higher [18]. This may be useful when considering how to address hotspots on the chip surface, but response times and chip workload need to be considered to understand if this is a practical approach.

The way in which flow in a microchannel is introduced to the chip surface can significantly change the efficiency of heat removal from the device. Flow impingement does this by forcing fluid through a small opening or nozzle as it sharply changes direction against the surface. This creates turbulence near the surface by increasing the liquid's velocity and mixing in the boundary layer, which enhances the heat transfer rate due to higher convective heat transfer coefficients. Design considerations include the momentum of the fluid, how the fluid spreads out from the source of impingement, dead zones in the flow, and overall pressure loss. One study concluded that it is difficult for impingement to address surface temperature fluctuations due to jet stagnation zones that develop in "dead space" of the nozzle, and the lack of ability to scale by simply adding additional nozzles [19]. It may be possible to overcome these challenges with silicon microchannel manifolds, as feature size control and impingement flow can be more precisely controlled with manifolded silicon microchannel features.

Manifold designs include multiple layers of processing, which increase cost and complexity, but address the challenging issue of fluid distribution to maintain minimal thermal cross-talk, pressure drop, and routing to hotspots. Co-designing the channel design with proper manifolding and channel dimensions can lead to decreased pressure drops by up to 29% and lower overall thermal resistances on the order of 20% [20].

B. Wafer Processing

The majority of approaches for using microchannel cooling have been at the chip level. It would be worthwhile to investigate taking advantage of standard wafer-level processes to implement silicon microchannels. A high-level process flow for putting a heat exchanger into direct contact with SOC and HBM chips was created as an illustration. The first view concerns the case in which the SoC and HBM have different heights (Fig. 3 and Fig. 4). In this case, silicon microchannels are etched into separate wafers, which are bonded to the SoC and HBM wafers, and then singulated, so that each chip has an attached silicon microchannel die bonded to it. This process could be done for each SoC and HBM die.

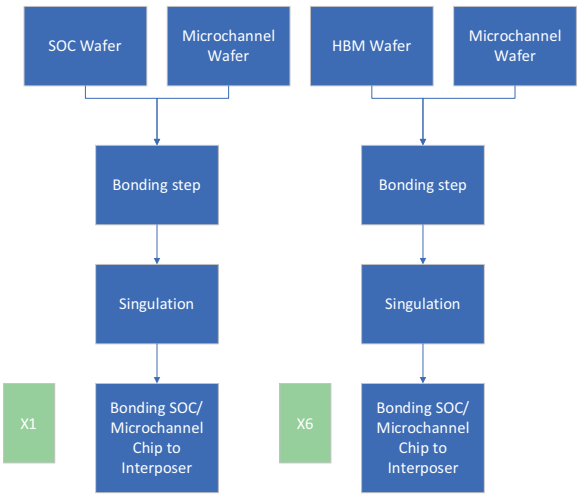

Fig. 3. Process flow for SoC and HBM stacks at different heights

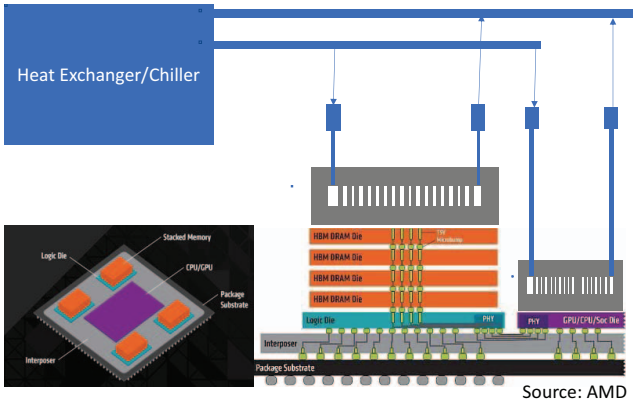

Fig. 4. Simplified view of vertical extent of SoC and HBM stack

The next approach is the case in which the SoC and HBM have similar heights. In this case, the silicon microchannel wafer is singulated and bonded to the entire chip package after the chiplets are bonded to the interposer (Fig. 5).

Industrial printing is a parallel industry approach to routing fluid though ICs that uses many of the same approaches being proposed for HPC microchannel cooling. HP applies silicon machining using laser ablation, wet etching, and DRIE to produce MEMS devices at high volume [21]. HP plans to commercialize inkjet printing integrating manifolded silicon microchannels directly into thermal inkjet device substrates, addressing many of the yield and manufacturing challenges that will also need to be solved by advanced heterogeneous packing. The processes needed to protect small MEMS features (<100 µm) are applicable to yield improvement and integration of silicon microchannels into the device layers. Fig. 6 shows a simplified process flow for MEMS devices used by HP. Cleanroom processing is necessary for joining and assembly of fluid delivery manifolds; a similar approach is required for manifolded silicon microchannel coolers.

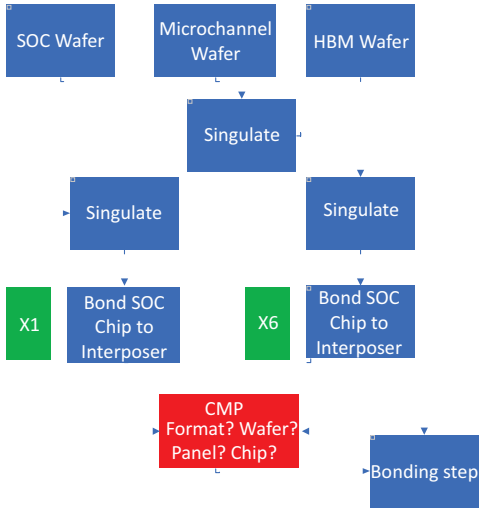

Fig. 5. Bosch process steps

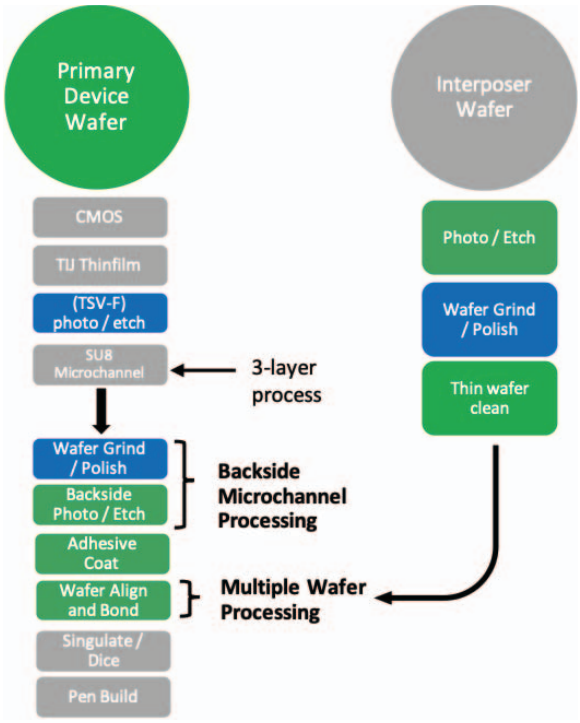

Fig. 6. HP's microchannel wafer process flow for industrial printheads

An interface is required between the small openings of the microchannels and the larger-diameter tubes that supply and drain the coolant. In 3D IC packages, cooling via both the top and the bottom of the chip stack may be used to enhance heat removal. One study designed a 3-board PCB system to provide both electric and fluidic interfaces to the chip [14]. Scaling this solution to heterogeneous chip package substrates with high interconnect density will require new innovations in substrate technology. Fig. 6 illustrates one such approach, which integrates with existing high-density board processes by selectively etching copper features to define fluidic channels with narrow line and space resolution.

Building complex packages involves considerations including material compatibility (e.g., survivability in coolants) and mechanical stability (e.g., CTE of materials, built in stress). Using poor thermal conductors to bond inner layers can significantly reduce the effectiveness of the performance. Even very thin oxide or adhesive layers can result in 30-50% reduction in performance due to the increase in thermal resistance. For this reason, it is important to develop reliable bonding methods, such as plasma bond, to ensure minimal thickness of a thermal interface material used, and to ensure high quality bonds formed with no chance for delamination leading to inner layer leaks. This is especially true for bottom of stack cooling, as electrical routing will be very close to fluidic routing.

C. Heat Transfer Design and Fluids

Systems utilizing microchannel cooling involve the use of fluids to transfer heat from the chip to an external cooling system. At a high level, key design choices involve whether the system will be a single-phase or two-phase (in which the cooling fluid is heated to the vapor phase and returns to the fluid phase when cooled), and type of the fluid to be used (generally, dielectrics are required due to the potential for shorting if the fluid were to contact the circuitry).

In single-phase liquid immersion cooling, electrical components, devices, and sub-systems are immersed in a heat transfer fluid that is either passively circulated by the natural convection of the heated coolant or actively circulated by pumping the coolant in, through and around the electrical device being cooled, and then transferring the heat absorbed by the coolant to a heat rejection device such as a radiator, dry cooler, liquid-to-liquid heat exchanger, or cooling tower. Propylene glycol and treated water heat transfer fluids are the predominant industry solution for single-phase liquid cold plates [22, 23]. To adapt these coolants for silicon microchannels, it will be necessary to add silicon to the wetted materials list and use compatibility testing to ensure that the material does not pose any reliability concerns.

Two-phase systems use pressure and phase-change of fluids to increase heat capacity. These fluids used in such systems undergo high pressures as they transition to a gas and then are re-condensed to remove heat. This can require complex systems with safety measures to ensure there is no build-up of pressure, and that the coolant does not boil off into the environment. Most two-phase cooling systems require that water be brought directly into the data room to chill the condenser and remove the collected heat. The boiling action of two-phase systems can create micro-cavitation, eroding metals in the electronics being cooled, as well as the metal components of the cooling system.

The most common two-phase dielectrics are fluorinated fluids which have relatively low boiling points and high density. The per-and polyfluoroalkyl substances (PFAS) are a group of chemicals used in dielectric fluids. These fluids generally have low freezing points and low viscosities at low temperatures, and are non-combustible. However, these fluids are very expensive,

and have poor thermal properties. Concerns about the continued use of such substances have emerged due the fact that they do not break down in the environment, can move through soils and contaminate drinking water sources, and bioaccumulate in fish and wildlife. At the end of 2022, 3M, a leading supplier of engineered fluids used in cooling, announced that it will discontinue production of PFAS-based materials by the end of 2025.

Some dielectrics used in industry are not appropriate replacements, as they can be toxic (aromatics, such as alkylated benzene, diethyl benzene, dibenzyl toluene, diaryl alkyl, and partially hydrogenated terphenyl), or hygroscopic and flammable (silicate-ester, known commercially as Coolanol 25R). Other dielectrics, including aliphatic hydrocarbons (polyalphaolefins, or PAOs) and synthetic polymeric compounds (dimethyl- and methyl phenyl-poly, or siloxane, commonly known as silicone oil) are non-toxic and non-volatile, and are used in some electronics applications such as cooling transformers, due to their high thermal conductivity and specific heat and low viscosity. However, these materials have low density and may not have high enough boiling points. Other materials that have been investigated for liquid coolants include nanofluids, phase-change materials, and liquid salts, but their use in electronics cooling is unproven.

V. Conclusions and Next Steps

The committee concluded that it is important to investigate wafer-level approaches to microchannel-based chip cooling. This involves the fabrication of microchannels on a separate wafer which is then integrated with individual die or an entire chip package. This approach minimizes processing challenges and reduces risk. The steps toward evaluating such an approach would involve defining a target test vehicle (the committee has developed a proposed test vehicle specification) and identifying steps needed to enter into prototype production. Care would need to be taken to mitigate any issues that could be barriers to production. The committee suggests that a collaborative approach be followed to evaluate this approach. Finally, the committee has identified next steps for commercialization of silicon microchannels used in HPC packages.

A. Potential Showstoppers and Barriers

Key issues considered include future-proofing development for IT roadmap; integrating silicon microchannels into SoC fabrication, SoC assembly and test, and server installation workflows, and coolant delivery systems challenges.

IT roadmaps for chip packages show substrate sizes increasing, becoming more challenging to manage. Any silicon microchannel process will need to scale accordingly to enable this roadmap. For example, if simple channels are etched directly into the top of the device substrate (SoC, HBM, etc.), the first manifold layer will need to be bonded to the SoC in such a way as to enable fine pitch integration (<100 μm), manage compressive and tensile stress induced from bow introduced during assembly of the SoC onto organic substrates, as well as from extreme temperatures experienced during shipping and operation.

Cold plates are currently assembled and yielded independent of the GPU, then brought together during the server manufacturing flow [24]. Due to the high cost of the SoC, it will be critical to ensure high quality microchannels can be formed and integrated as late in the process as is reasonable to minimize cost impact from yield. The current process flow calls for cold plates to be assembled, leak tested, filled with coolant, and then have the coolant drained before shipping for server installation. Managing coolant priming and start-up issues may be challenging in smaller capillaries, especially if internal channels must be drained/dried and the subsequent challenges to prime and ensure no trapped air inside the channels.

In order to achieve higher performance by reducing microchannels to less than 100 μm careful manifolding is required to maintain reasonable pressure drops. A range of typical node inlet to exit pressure drops that in row coolant distribution units (CDUs) can achieve < 250 kPa @ 1-2 LPM [25]. Thermal resistance and flow impedance directly trade-off against each other at higher flow rates. Given that this is the budget for the entire system, a reasonable target for the silicon microchannel portion of the budget needs to be < 100 kPa. Although not impossible, this will require more complex layers of manifolding than typically demonstrated in literature to deliver on high flow rates at low pressure drops.

Key issues to consider in integrating microchannel-based chip cooling include:

- Industry and supply chain needs to build confidence in silicon microchannels to overcome perceived inherent reliability risk (e.g., liquids integrated with electronics)
- An already complex list of wetted materials for coolants needs to have silicon added to it in order to address reliability in concerns
- Typical CDUs are designed today to support 50 μm particle filtration levels. As silicon microchannels scale to < 100 μm, it will be necessary to improve filtration levels accordingly.

B. Identification of Steps Needed to Put into Production

To address the potential technology showstoppers, an analysis of the full system payback and total cost of ownership needs to be done comparing state-of-the-art liquid cooled solutions. In order to do this, it will be necessary to work with the data center industry community on the approach to address the main requirements needed to deliver silicon microchannel cooling solutions:

- SoC package design and integration
- SoC manufacturing test and assembly
- Coolant vendors wettable materials list and reliability
- Cooling solution providers integrated components such as in-row and rack coolant distribution units that can provide the right filtration, flow rate, and pressure drops for narrow microchannels.
- Total cost of ownership of silicon microchannel cold plates vs conventional

Currently, there are no dedicated EDA tools for designing the layout for HPC chip packages with cooling. Tools to resolve routing conflicts between cooling channels, power lines, and vias are needed for optimal integration of microchannels.

Visibility into changes of materials, architectures, and manufacturing approaches for interposers in HPC packages is needed in order to plan future microchannel cooling implementations.

C. Next Steps: Call to Action for Si-Microchannels Phase II

The committee recommends two follow-on activities. First, the development of microchannel-based chip cooling approaches should be pprioritized by the CHIPS program. The CHIPS Act authorizes the creation of the National Advanced Packaging Manufacturing Program (NAPMP), which the CHIPS program expects to fund an Advanced Packaging Pilot Facility, a pilot packaging facility to test and pilot new equipment and processes. This would be an ideal facility to evaluate prototype microchannel-based cooling approaches.

Second, the committee recommends a Si Microchannels Phase II program with the goal of creating test vehicle(s). This would involve forming an industry team, including SEMI, to create one or more test vehicles to demonstrate Si microchannel chip cooling, as outlined in this report. This effort would coordinate with industry and research partners to address technical and industry issues highlighted in this paper.

ACKNOWLEDGMENT

The authors acknowledge the support of Melissa Grupen-Shemansky, Vice President, SEMI.

REFERENCES

[1] J. H. Lau, "Recent advances and trends in advanced packaging," in IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 12, no. 2, Feb. 2022, pp. 228-252.

[2] A. Bar-Cohen, "Gen 3 “embedded” cooling: key enabler for energy efficient data centers," IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 7, no. 8, Aug. 2017, pp. 1206-1211.

[3] J. Hung, et al., “Advanced system integration for high performance computing with liquid cooling,” 2021 IEEE 71st Electronic Components and Technology Conference (ECTC).

[4] K.P. Drummond, et al., “Characterization of hierarchical manifold microchannel heat sink arrays under simultaneous background and hotspot heating conditions,” International Journal of Heat and Mass Transfer, Volume 126, Part A, 2018, pp. 1289-1301.

[5] K.W. Jung, et al, “Embedded cooling with 3D manifold for vehicle power electronics application: single-phase thermal-fluid performance,” International Journal of Heat and Mass Transfer, Volume 130, March 2019, pp. 1108-1119.

[6] T.E. Sarvey, et al., “Microfluidic Cooling of a 14-nm 2.5-D FPGA with 3-D printed manifolds for high-density computing: design considerations, fabrication, and electrical characterization,” IEEE Transactions on Components, Packaging and Manufacturing Technology, Vol. 9, No. 12, December 2019, pp. 2393-2403.

[7] X. Zhang, Y. Han, G. Tang, H. Chen, and B.L. Lau, “Development of advanced liquid cooling solution on data centre cooling,” 2022 IEEE 72nd Electronic Components and Technology Conference (ECTC).

[8] M. Bergendahl, et al., "Stacked silicon microcoolers," 19th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2020, pp. 69-76.

[9] T-W Wei, et al., “Demonstration of package level 3D-printed direct jet impingement cooling applied to high power, large die applications,” IEEE 70th Electronic Components and Technology Conference (ECTC) , 2020, pp. 1422-1429.

[10] D. B. Tuckerman and R. F. W. Pease, "High-performance heat sinking for VLSI," IEEE Electron Device Letters, vol. 2, no. 5, May 1981, pp. 126-129.

[11] R.J. Phillips, “Microchannel heat sinks,” The Lincoln Laboratory Journal, Volume I, Number 1, 1988, pp. 31-48.

[12] M. D. Schultz, et al., "Microfluidic two-phase cooling of a high-power microprocessor part A: Design and fabrication," 2017 16th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), Orlando, FL, 2017, pp. 450-457.

[13] T. E. Sarvey, et al., "Embedded cooling technologies for densely integrated electronic systems," 2015 IEEE Custom Integrated Circuits Conference (CICC), San Jose, CA, 2015, pp. 1-8.; T. E. Sarvey et al., "Monolithic Integration of a Micropin-Fin Heat Sink in a 28-nm FPGA," IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 7, no. 10, Oct. 2017, pp. 1617-1624.

[14] R. van Erp, G. Kampitsis, L. Nela, R.S. Ardebili, and E. Matioli., “Embedded microchannel cooling for high power-density GaN-on-Si power integrated circuits,” 19th IEEE ITHERM Conference, pp. 53-59, 2020; R. van Erp, R. Soleimanzadeh, L. Nela, G. Kampitsis, and E. Matioli, “Co-designing Electronics with Microfluidics for more Sustainable Cooling,” *Nature,* vol. 585, 211.

[15] S. Hazra, et al., “Parametric design analysis of a multi-level 3D manifolded microchannel cooler via reduced order numerical modeling,” *International Journal of Heat and Mass Transfer*, Volume 197, 2022.

[16] G.S. Lyon, Fluid heat exchanger configured to provide a split flow, US Patent 9603284B2, filed May 20, 2014, issued Mar. 21, 2017.

[17] H-C Chiu, M-J Youh, R-H Hsieh, J-H Jang, and B. Kumar, “Numerical investigation on the temperature uniformity of micro-pin-fin heat sinks with variable density arrangement,” Case Studies in Thermal Engineering, Volume 44, 2023.

[18] X. Li and Y. Xuan, "Embedded cooling of high-heat-flux hotspots using self-adaptive microchannel/pin-fin hybrid heat sink," IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 12, no. 10, Oct. 2022, pp. 1627-1635.

[19] Z. Zhang, X. Wang, and Y. Yan, “A review of the state-of-the-art in electronic cooling,” e-Prime - Advances in Electrical Engineering, Electronics and Energy, Volume 1, 2021, 100009.

[20] C. Chen, X. Wang, B. Yuan, W. Du, and G. Xin, “Investigation of flow and heat transfer performance of the manifold microchannel with different manifold arrangements,” Case Studies in Thermal Engineering, Volume 34, 2022.

[21] J. Stasiak, S. Richards, and P. Benning, “Hewlett-Packard’s MEMS technology: thermal inkjet printing and beyond,” Microelectronics to Nanotechnologies: Materials, Devices & Manufacturability, edited by A.B. Kaul, Boca Raton: CRC Press, 2012.

[22] “Guidelines for using propylene glycol based heat transfer fluids in single phase cold plate-based liquid cooled racks,” Open Compute Project.

[23] “Guidelines for using water-based transfer fluids in single-phase cold plate-based liquid-cooled racks,” Open Compute Project.

[24] “Liquid cooling integration and logistics white paper,” Open Compute Project.

[25] “An Advanced liquid cooling rack design for data center,” Open Compute Project.