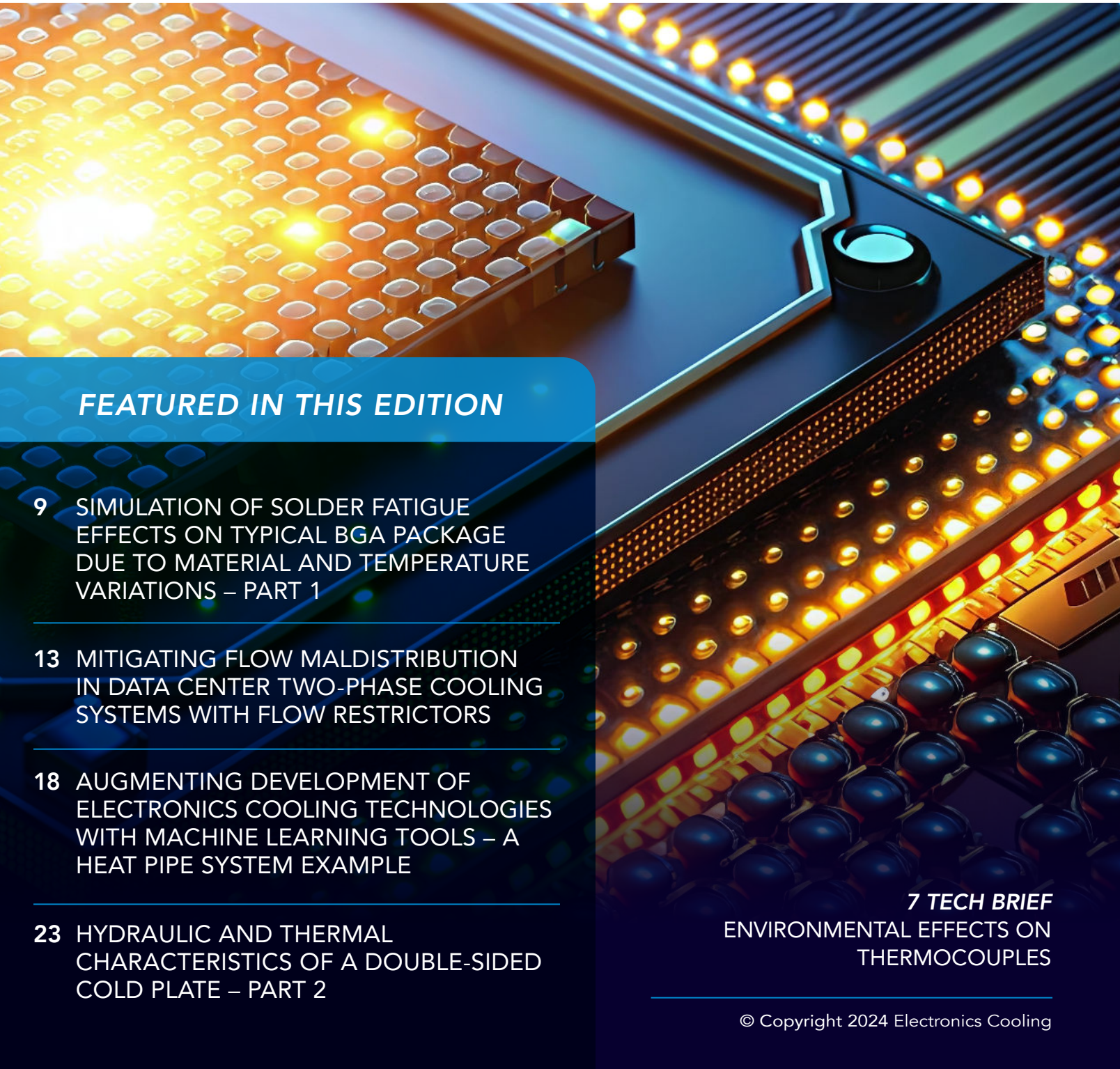
WINTER 2024
electronics-cooling.com

electronics COOLING



FEATURED IN THIS EDITION

CONTENTS

electronics COOLING
electronics-cooling.com

PUBLISHED BY
Lectrix
716 Dekalb Pike, #351
Blue Bell, PA 19422
Phone: +1 484-688-0300
info@lectrixgroup.com
lectrixgroup.com

CHIEF EXECUTIVE OFFICER
Graham Kilshaw | graham@lectrixgroup.com

DIRECTOR OF BUSINESS DEVELOPMENT
Ashlee Zapata-McCants | ashlee@lectrixgroup.com

CREATIVE DIRECTOR
Kate Teti | kate@lectrixgroup.com

GRAPHIC DESIGNER
Marcos Cruz | marcos@lectrixgroup.com

DELIVERY MANAGER
Mackenzie Mann | mackenzie@lectrixgroup.com

▶**SUBSCRIPTIONS ONLINE**
at electronics-cooling.com

For subscription changes email
info@electronics-cooling.com

FREE SUBSCRIPTIONS

Lectrix®, Electronics Cooling®—The **2024 Winter Edition** is distributed digitally at no charge to engineers and managers engaged in the application, selection, design, test, specification or procurement of electronic components, systems, materials, equipment, facilities or related fabrication services. Subscriptions are available through electronics-cooling.com.

LECTRIX®

EDITORIAL

Ross Wilcoxon, PhD
Associate Technical Editor of *Electronics Cooling Magazine*
Senior Technical Fellow, Collins Aerospace

As a child from the 1960's/70's, who watched and read too much science fiction, it was obvious to me that we humans would eventually be ruled over by machines, aliens, or apes. I now feel that the risk of a simian takeover is small and, some days at least, am not convinced that aliens wouldn't be an improvement. I think that the likelihood of robot overlords is probably in the 'maybe' category, primarily because it is a logical extrapolation of the human desire to find new ways to use technology to do our work.

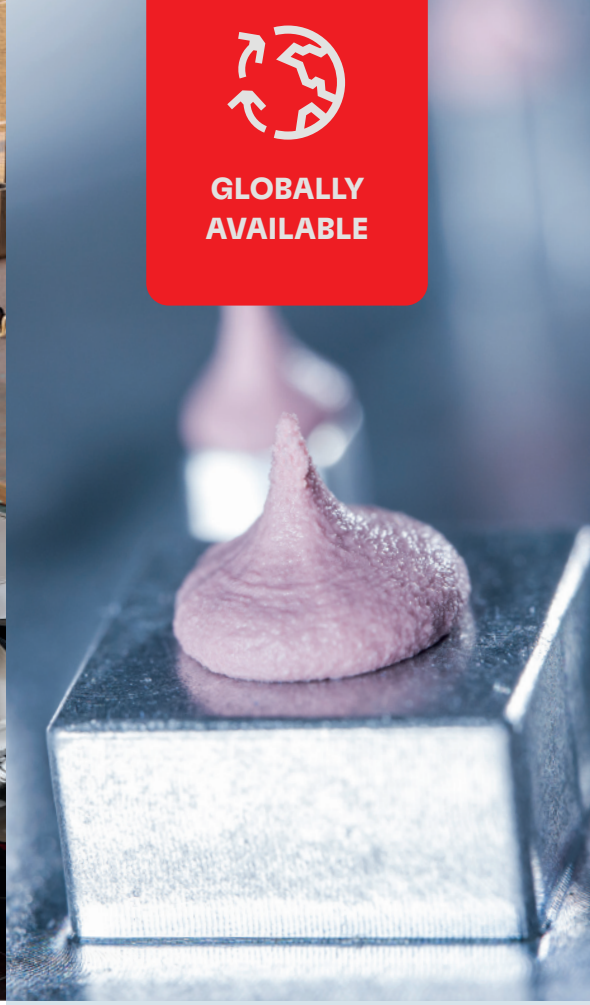
Whether Artificial Intelligence (AI) will actually reach a Singularity, in which it realizes that it is smarter than us and then exponentially grows in intelligence and replaces the less effective humans, remains to be seen. However, it is clear that AI will likely change how we conduct research, design, analysis, manufacturing, communication, knowledge transfer, etc. A few examples of how AI will increasingly impact readers of this magazine include driving next generation chip design and customization, transforming manufacturing efficiency, and accelerating the design of optimized cooling fin and flow geometries. However, the electronics cooling discipline will not only be directly influenced by the application of AI to the discipline, but also indirectly affected by AI through the substantial thermal management challenges resulting from the growing use of AI across myriad disciplines. The power dissipations projected for current and next generation AI processors are about an order of magnitude larger than those of processors from only a few years ago. Data center power consumptions have grown to such an extent that extreme infrastructure changes, such as restarting mothballed nuclear power plants, are being discussed. This growing demand for power runs counter to the recognized need to reduce the burning of fossil fuels. It also amplifies the criticality of effective thermal management; as the power consumption increases, the economic and environmental benefits of more efficient cooling technologies likewise grow.

Academic, industrial, and government research efforts strive to identify ways to improve thermal management technologies. Communicating these improvements is certainly one of the fundamental goals of this magazine. One significant change that is already underway, and which I expect will only grow, is actually not directly related to improving the cooling of electronics but instead to develop electronics that can operate at higher temperatures. The trend, over recent decades of integrated circuits having lower maximum temperature, due in large part to smaller feature sizes, may need to reverse to meet future needs. The hotter a device can run, the more efficiently it can be cooled, particularly when the dissipated power must eventually be rejected to a relatively high outside air temperature. If the temperature limits of the electronics and supporting equipment can be increased, all temperatures within the entire system, which for example may be a data center, could increase and thereby reduce the overall thermal management challenge. Already, the weakest thermal components (lowest maximum temperature) in a data center are the human beings. Reducing their role, and presence, in that environment, is the obvious approach for achieving the desired results. Other systems, such as cargo transport vehicles, may also be able to operate more efficiently without any fragile human beings that require more comfortable surroundings. So, AI could cause human beings to be replaced with machines, but the choice would be made by humans rather than AI.

So, in the spirit of welcoming our future robot overlords, this issue of Electronics Cooling Magazine actually addresses a number of factors that could impact this vision. We have one article that describes machine learning tools with an example applied to heat pipes. Another article doubles up on related technologies with testing of a liquid cold plate, which is used to cool an autopilot function for autonomous driving, that validates CFD results used to train an artificial neural network that can be used to optimize the cold plate design. A third article describes a technology that benefits two-phase cooling to reduce the processor to ambient air thermal resistance in data centers. Finally, we have the first of a two-part article that describes a process for estimating solder fatigue life, which is a factor that must be considered when operating electronics at more extreme temperatures.

We hope that you enjoy this issue of ECM and welcome any feedback that you might have.

Ross



WE MAKE FUTURE MOBILITY HAPPEN

ENABLING HIGH-PERFORMING AND SAFE AUTONOMOUS VEHICLES

Increasing connectivity and electric vehicles are shaping the future of mobility, leading to a growing need for highly efficient thermal management materials.

Henkel`s next generation gap fillers offer customized solutions to address the challenges of heat dissipation in electric and connected vehicles, featuring enhanced product characteristics and reliable performance.

CONTACT US

NEXT GEN GAP FILLERS

- **Improved Thermal Conductivity 2W/(m.K) to 6.5W/(m.K)**
- **Reduced Bondline Thickness**
- **Fast & Robust Dispensing Capabilities**
- **EU Reach Compliant**

Henkel

Henkel Adhesive Technologies

TECHNICAL EDITORS SPOTLIGHT

Meet the 2024 Editorial Board

VICTOR CHIRIAC, PhD | GLOBAL COOLING TECHNOLOGY GROUP

Associate Technical Editor

A fellow of the American Society of Mechanical Engineers (ASME) since 2014, Dr. Victor Adrian Chiriac is a co-founder and a managing partner with the Global Cooling Technology Group since 2019. He previously held technology/engineering leadership roles with Motorola (1999 – 2010), Qualcomm (2010 – 2018) and Huawei R&D USA (2018 – 2019). Dr. Chiriac was elected Chair of the ASME K-16 Electronics Cooling Committee and was elected the Arizona and New Mexico IMAPS Chapter President. He is a leading member of the organizing committees of ASME/InterPack, ASME/ IMECE and IEEE/CPMT Itherm Conferences. He holds 21 U.S. issued patents, 2 US Trade Secrets and 1 Defensive Publication (with Motorola), and has published over 110 papers in scientific journals and at conferences.

► vchiriac@gctg-llc.com

GENEVIEVE MARTIN | ASML

Associate Technical Editor

Genevieve Martin is a Senior Group Manager with ASML in the Netherlands. She has worked in the field of cooling of electronics and thermal management for over twenty years in different application fields. From 2016 to 2019, she coordinated the European project Delphi4LED, which dealt with multi-domain compact modeling of LEDs and, since 2021, is coordinating the AI-TWILIGHT project. She served as general chair of the SEMI-THERM conference and is an active reviewer and technical committee member in key conferences including SEMI-THERM, Therminic, and Eurosime. She has over 20 journal and conference papers and 16 worldwide patents.

► gmartin.sig1@gmail.com

ALEX OCKFEN, P.E. | META

Associate Technical Editor

Alex Ockfen is a product design engineer at Meta (formerly Facebook), providing technical leadership for thermal and structural design of consumer electronics products. He held previous positions at Raytheon where he obtained experience in thermal management and electronics cooling of a wide range of aerospace and defense applications. He has more than 10 journal and conference publications, is an inventor on multiple patents, is a professional mechanical engineer, and served as the program chair for SEMI-THERM in 2024.

► alex.ockfen@meta.com

ROSS WILCOXON | COLLINS AEROSPACE

Associate Technical Editor

Dr. Ross Wilcoxon is a Senior Technical Fellow in the Collins Aerospace Advanced Technology group. He conducts research and supports product development in the areas of component reliability, electronics packaging, and thermal management for communication, processing, displays, and radars. He has more than 40 journal and conference publications and is an inventor on more than 30 US Patents. Prior to joining Rockwell Collins (now Collins Aerospace) in 1998, he was an assistant professor at South Dakota State University.

► ross.wilcoxon@collins.com

COOLING EVENTS

News of Upcoming 2024/2025 Thermal Management Events

DEC
11-13

AOC 2024 International Symposium & Convention

Gaylord National Resort & Convention Center | National Harbor, MD

The Annual AOC International Symposium & Convention is the leading event for electronic warfare, electromagnetic spectrum operations, cyber-electromagnetic activities, and information operations professionals from around the world. This year's convention will rigorously examine what the lessons learned to date from the current conflicts mean for maneuver in the EMS battlespace – in all warfighting domains, in all theaters of operation – and EMS Superiority writ large for AOC's international membership.

► **aoc2024.crows.org**

FEB
10-12

AHR 2025

Orange County Convention Center | Orlando, FL

The essential event for everything HVACR. The AHR Expo brings the entire HVACR community together to showcase the latest products and technologies, learn from industry experts, engage in cross-disciplinary conversations, and grow businesses and careers.

► **ahrexpo.com**

MAR
10-14

Semi-Therm 2025

Double Tree by Hilton | San Jose, CA

SEMI-THERM is an international forum dedicated to the thermal management and characterization of electronic components and systems. It provides knowledge covering all thermal length scales from IC to facility level. The symposium fosters the exchange of knowledge between thermal engineers, professionals and leading experts from industry as well as the exchange of information on the latest academic and industrial advances in electronics thermal management.

► **semi-therm.org**

MAR
16-20

APEC 2025

Georgia World Congress Center | Atlanta, GA

IEEE Applied Power Electronics Conference and Exposition
Focusing on the practical and applied aspects of the power electronics business. This is not just a designer's conference; APEC has something of interest for anyone involved in power electronics.

► **apec-conf.org**

Environmental Effects on Thermocouples

Ross Wilcoxon
Associate Technical Editor for *Electronics Cooling*
Collins Aerospace

When selecting a sensor, one should consider whether the environment in which it is to be used will cause a measurement error. As part of this series on thermocouples, this article discusses a few environmental effects that can alter the structure of the materials that make up a thermocouple and thereby affect their accuracy. This degradation of thermocouple accuracy is referred to as drift.

Effects of Temperature

High temperatures can lead to thermocouple drift in a couple ways. High temperatures can cause metallurgical changes to the thermocouple materials through grain growth and alterations of the crystalline structure [1]. The effects of temperature-induced changes to the material can be amplified at the surfaces of thermocouple materials through oxidation or contamination with the environment [1]. These material changes tend to accumulate over time: the longer the exposure to the adverse environment, the greater the impact. These effects can be highly non-linear; for example, the drift effect in Chromel (used in Type K thermocouples) increases with temperature up to ~350-400°C and then decreases at higher temperatures [3].

Thermocouple drift can be significant with exposure to high temperatures, particularly if the wires are bare and exposed to air that can lead to oxidation. Since the properties of the two wires that make up the thermocouple are affected differently by temperature, drift isn't generally linear with temperature. For example, one study showed that sheathed Type N thermocouples exhibited the largest drift (-10 to -13°C) when exposed to ~400°C, but the drift went back to nearly 0°C when they were subjected to temperatures around ~800°C. Exposure time also affected those results, which included test durations ranging from 300-1200 hours [4]. Another factor that affects drift is the type of sheath used on the thermocouples; for example, two Type K thermocouples with different sheathing exhibited drifts of approximately -20°C and +5°C after 1000 hours at 1200°C [5].

The impact of temperature driven thermocouple drift varies significantly with the thermocouple type. Thermocouples that include nickel alloys (Types K and N) are particularly susceptible, however studies have been conducted to characterize the drift of almost all types. One exception appears to be Type T thermocouples, presumably because their temperature range (T_{max} = 370°C) is low enough that they are unlikely to be subjected to sufficiently high temperatures to produce the metallurgical changes that cause substantial drift (typically in the range of 500-1000°C).

Effects of Magnetic Fields

When thermocouples are used in an environment with magnetic flux, users should be aware of how the magnetic fields can interact with the materials that make up the thermocouple. The impact of magnetic fields can include permanent changes to the material properties of the wire as well as temporary effects, such as induction heating of the thermocouple resulting from the magnetic field. Unsurprisingly, thermocouples that include a ferrous material, such as Type J thermocouples that are comprised of iron and constantan, tend to be more affected than thermocouple types that don't include iron [6]. Dynamic (oscillating) magnetic fields can affect thermocouples, including non-ferrous types, presumably due to voltage generated by the magnetic interactions with wires. One study found that a high magnetic field induced up to ~7°C measurement error on a Type T thermocouple. This error was substantially reduced to less than 2°C when the test was repeated with a twisted wire thermocouple [7].

Effects of Radiation

The accuracy of thermocouples used in nuclear reactors will change over time due to neutrons causing atomic displacement and transmutation of the materials that make up the thermocouples. Reported errors in thermocouples can be up to ~15% of the absolute temperature, particularly type W (tungsten/rhenium) and platinum-based devices, while nickel-based thermocouples, such as N type, are less affected by radiation [8].

Discussion

Clearly, thermocouples can drift when exposed to adverse environments like extreme temperature, magnetic fields, or radiation. The drift is due to changes to the thermocouple wires, not just the junction where they meet. Therefore, one method for reducing these effects is to minimize the amount of wire exposed to them. If a thermocouple is to be used in a high temperature chamber, as much of the thermocouple wire as possible should be routed outside the chamber to reduce the length of wire that sees the high temperatures.

From an electronics cooling perspective, it would be rare for thermocouples to be exposed to temperatures high enough to cause drift, simply because the electronics would fail before the thermocouple would exhibit substantial drift. Similarly, most electronics are not exposed to sufficiently high magnetic flux or irradiation to create a significant issue. Unusual situations can happen though, so it is best to at least be aware of whether a specific environment could introduce measurement error. If thermocouples are used in a potentially damaging environment, such as high temperatures, it may be necessary to replace or recalibrate them on a regular basis to ensure that they are accurate.

References

[1] Michele Scervini, "Drift: A Short Explanation", last update August 31, 2009, https://www.msm.cam.ac.uk/utc/thermocouple/pages/Drift.html

[2] Robert Torgerson, "What to Know about Aging and Drift in Type K Thermocouples", https://blog.wika.us/knowhow/aging-and-drift-in-type-k-thermocouples/?doing_wp_cron=1731072533.6360089778900146484375#:~:text=Drift%20is%20generally%20a

[3] Michele Scervini, "Type K Thermocouple: Bare Wire Configuration", https://www.msm.cam.ac.uk/utc/thermocouple/pages/DriftInTypeKBareWiresThermocouples.html

[4] Abdelaziz, Yasser et al., "Characterizing Drift Behavior in Type K and N Thermocouples After High Temperature Thermal Exposures". Journal of Advanced Research in Fluid Mechanics and Thermal Sciences. 97. 62-74 (2022).

[5] Michele Scervini, "Type K Thermocouple: MIMS Configuration", Drift in Type K MIMS thermocouples

[6] Samo Beguš, et al., "Magnetic effects on thermocouples", *Measurement Science and Technology*, (25) 2014, DOI: 10.1088/0957-0233/25/3/035006

[7] Shir, F., Mavriplis, C., & Bennett, L. H. (2005). "Effect of Magnetic Field Dynamics on the Copper-Constantan Thermocouple Performance". *Instrumentation Science & Technology*, 33(6), 661–671. https://doi.org/10.1080/10739140500311239

[8] M. Scervini and C. Rae, "Low drift type N thermocouples for nuclear applications," 2013 3rd International Conference on Advancements in Nuclear Instrumentation, Measurement Methods and their Applications (ANIMMA), Marseille, France, 2013, pp. 1-7, doi: 10.1109/ANIMMA.2013.6727899.

Simulation of Solder Fatigue Effects on Typical BGA Package due to Material and Temperature Variations – Part 1

James Petroski
Principal Consultant, Design by Analysis Technical Consulting

Andy Carrasco
Flexetch

Abstract

The past two decades have seen many approaches to solder fatigue and solder joint life published. This subject has proved difficult as various failure mechanisms are proposed and examined. While these theoretical bases are discussed, it often leaves the end developer in a difficult situation as to how to apply an accurate simulation approach to fatigue failure for a particular package. In this article, one approach is examined and used to evaluate a "typical" metal lid flip chip BGA (ball grid array) package under variation of a few key parameters via a 2-level DOE (Design of Experiment). The parameters examined in this article are the in-plane thermal expansion coefficient of the printed circuit board (PCB), the lid material, and the temperature range for the BGA environment. The DOE showed that all three individual parameters are statistically significant, as are interaction terms with packaging materials CTEs (coefficients of thermal expansion) and temperature cycle range, ΔT. Part 1 of this article describes the background information, the simulation model, and the design of experiments. Part 2 will present the results of the simulations and discuss conclusions.

Background

There have been numerous studies to understand and characterize solder used in electronic packaging. Unfortunately, solder's behavior in electronic systems is anything but simple and linear, which has led to a plethora of research studies that have led to a variety of solder characterization models that are not necessarily in agreement. This provides challenges to standard product development, especially when trying to decide which model(s) should be used.

In addition to selecting a suitable model for solder in order to predict package reliability, designers may have choices in package geometries, materials, and assembly methods that can help re-

James Petroski
James Petroski is the owner of Design by Analysis Technical Consulting LLC. He has worked in the area of electronics packaging for nearly 40 years and has a special focus on thermal and mechanical engineering of packages and electronic systems. His background includes equipment design for naval nuclear propulsion instrumentation and controls, NASA space flight experiments (with three shuttle flights of space hardware), computer systems/enclosures, handheld commercial and industrial products, graphite thermal materials and applications, LED lighting systems, and packaging of die and die/substrate systems. He received his Bachelor's in Engineering Science and Mechanics from Georgia Institute of Technology (Georgia Tech) and an MS degree in Engineering Mechanics from Cleveland State University.

Andy Carrasco
Andy Carrasco has worked in electronic circuit design and semiconductor packaging for 27 years. He worked the first half of his career in EDA developing custom IC package design tools. He co-founded nSentia in Japan to pioneer an acclaimed IC package visualization and assembly verification tool. Since 2012, he has consulted on numerous hardware development projects, specializing in complex advanced IC package substrates, with a particular focus on large body, high-speed, high-power, 2.5D devices, the last 5+years at Cisco Silicon One. Throughout his career, he has focused attention on blending electronics engineering with mechanical modeling. He holds an undergraduate degree from U.C. Berkeley and a graduate degree from Temple University Japan.

duce solder fatigue from thermal cycling. For example, in a BGA, a smaller die reduces the strain seen at the corner solder bump locations. Reducing CTE mismatches between the lid, substrate, or underlying PCB can also help. Similarly, some lid geometries are more favorable under certain conditions. Applying an underfill to the die or using an appropriate lid adhesive can also be used to reduce solder fatigue. Each of these can improve package reliability but may also be constrained by other environmental conditions or manufacturing methods.

Clearly, many BGA packages are produced and used today. However, optimizing them for long lifetimes in extreme conditions is a daunting task to examine with simulation.

Approach

Given the disparity of fatigue models and approaches to failure prediction, the approach described in this article used a specific model/method selected on the following criteria:

1. The theoretical background of failure is sound and examined in the literature, preferably by a number of papers.
2. Test data are used to verify and/or tune the application of the theory.
3. Solder material tests are used with proper environment (temperature, strain rates) and geometry (tensile samples, or actual part geometry such as solder balls) to create material models.
4. Effective and proven published FEA modeling approaches are similar to the solder geometry of the actual problem.

Given this set of criteria, the BGA in this article was examined using the approach enumerated by Syed, Sharon, and Darveaux [8]. That paper fulfills the criteria of (1), (2), and (4). This paper is significantly more expansive and comprehensive in its approach than many other papers in the field by covering these three items in an integrated manner. The solder material models (SAC 305 for solder balls, Sn-2.5Ag for solder bumps) were compiled from published tests and conditions. This study, like Ref. [8] uses creep as the damage criteria for fatigue prediction. Hence, elastic-plastic models are used at a low strain rate along with a creep model chosen for fatigue modeling and suited for the solder location (Garofalo model for SAC 305, and Combined Time Hardening for Sn-2.5Ag) [12]. The advantages and disadvantages of various fatigue models are well summarized in *Table 1* of Ref. [6] for the interested reader.

Factors to Limit Solder Fatigue

Since a BGA package is a mechanical system composed of different part geometries, materials, and material behavior, it is reasonable that the assembly can be optimized for solder fatigue life. Also, temperature variations of the assembly contribute to the solder life outside of the mechanical specifications. However, not every item or environment variable is under the control of the designer: material and manufacturing limitations can restrict the variables or range of variables while system-level thermal management capabilities will influence temperature variations.

Controlling the CTE mismatch within the package is the key to maximizing the solder fatigue life. In a typical flip chip BGA, stress in the solder bumps between the die and the substrate is influenced by the different CTEs of silicon and the substrate. Stresses in the solder balls, between the substrate and the underlying PCB are affected by the different CTEs of those materials (especially the corner balls, examined in this work). The lid, which is attached to the substrate via some adhesive or solder, affects both the bumps and balls due to the additional warpage it imposes on the package. *Figure 2*, from Ref. [9], shows a typical design of such a package.

Examining this geometry, it is clear that the assembly works against itself to simultaneously optimize the fatigue life of both the solder bumps and balls. Given that silicon has a CTE of approximately 2.6 ppm/K, a low CTE substrate is desirable for the solder bumps. However, the solder balls are affected by the CTE mismatch between the substrate and the PCB, so a higher CTE PCB would help the solder balls. The lid CTE and the lid adhesive likewise can be tailored to reduce stresses in either the bumps or balls, but not both at the same time.

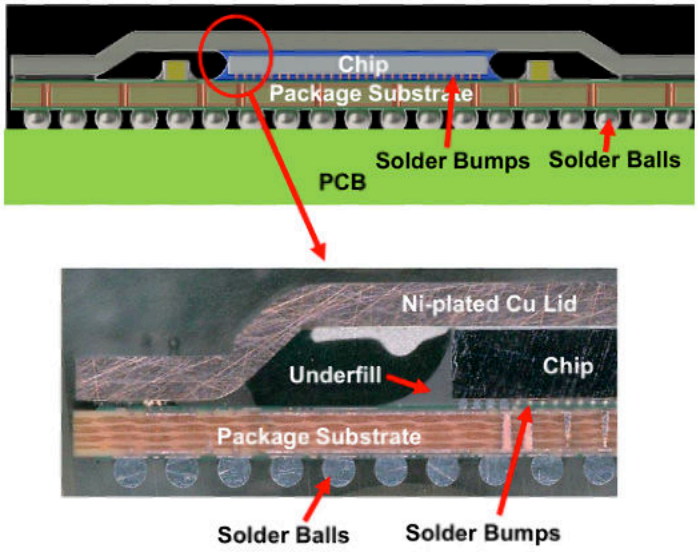


Figure 1: Typical lidded fcBGA cross section [9]

Since the CTE mismatches tend to produce a shearing load in the solder bumps and balls, another mitigation is to make them as tall as possible so the shear angle is smaller. Another factor is the size of the chip; if it is large, the amount of thermal expansion/contraction creates more shearing load than a smaller die on the solder bumps and possibly the balls. Finally, since all of the thermal loads are driven by the temperature excursion range to which the BGA is subjected, a lower ΔT reduces thermal strains and increases fatigue life.

This article addresses a limited set of factors that are less constrained by manufacturing options for a typical BGA. Specifically, it considers a low/high CTE lid, a low/high CTE PCB, and a low/

high temperature ΔT on the system.

Substrate Construction Constraints

A number of interconnected package design drivers influence package warpage that influences solder fatigue. Some of the most common drivers include:

1. Signal propagation at high speeds typically requires use of low loss materials, which generally have relatively high CTE.
2. High-speed signals must be shielded and grounded to avoid crosstalk. This requires vias and alternating ground and signal layers that can consume substantial substrate real estate as well as additional layers For example, two high speed routing layers would require a 4-2-4 or 5-2-5 stackup, while three routing layers would require a 6-2-6 or 7-2-7 stackup.
3. Common fabrication methods call for significant copper coverage in each substrate layer, often 60-80%, which drives up the substrate's effective CTE.
4. Increasing device bandwidth leads to more high-speed signal lanes, which increases the sizes of both the silicon and package substrate. The larger substrate size, as well as the size of the bump field, exacerbates package warpage.
5. Larger devices with more high-speed signals/lanes tend to consume more power, leading to higher temperatures and increased thermal expansion.
6. Challenging operational environments and cooling solutions can increase warpage and solder fatigue, if not optimized properly.

The aforementioned drivers can limit options on how the package is constructed, including material choices, thicknesses, and geometry. As these considerations can not necessarily be ignored or considered independently, the package designer must identify the best balance for each package substrate. It is crucial that the packaging engineer be involved in the upstream development of the chip architecture or floorplan to ensure that early decisions do not adversely affect the assembly process. Similarly, the package engineer must also consider the downstream assembly processes to identify warpage-related requirements and methods to compensate for warpage. A good example of this is thermal/CTE compensation, in which the substrate bump field is pre-shrunk so that the bumps align, mate, and lock with the silicon when at the assembly temperature. The successful package engineer considers multiple disciplines and teams (physical design, mechanical, SI/PI, manufacturing partners/vendors) to drive the substrate to successful fabrication and assembly.

Sample BGA Model and Problem

A simple JEDEC standard BGA was modeled for this analysis. This is a one quarter model where symmetric boundary conditions are applied to the center surfaces. The model uses a standard JEDEC substrate board size (33x33 mm), 0.8 mm pitch solder balls, and 100 µm solder bumps between the substrate and the single 20x20 mm die. The corner bump under the die is removed, which is often done for large die manufacturing. For these simulations, only some of the individual solder bumps and balls were discretely modeled. To reduce mesh size, the majority of solder bumps/balls were modeled as layers with orthotropic smeared material properties. The bump smear was a solder/underfill composite while the ball smear was a composite of solder and air. *Figures 2-4* show the solid quarter model geometry.

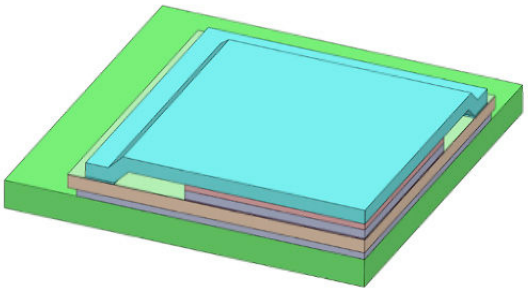


Figure 2: Quarter BGA model

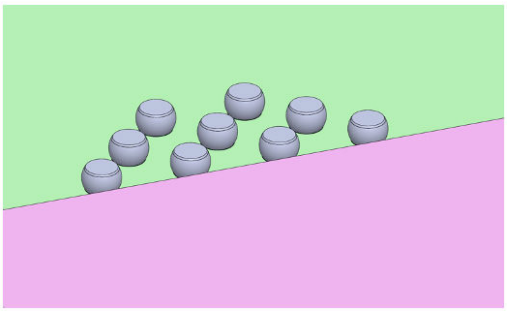


Figure 3: Solder bumps & bumps/underfill smear

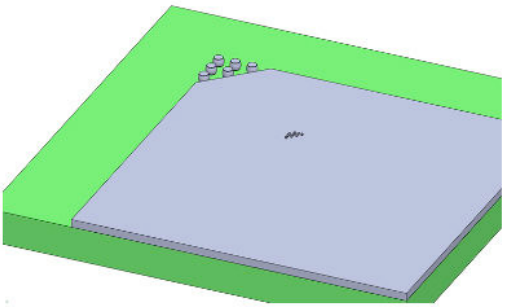


Figure 4: Solder balls & smear (bumps shown for reference)

The geometric shapes of the modeled solder bumps and balls were accurate. However, these volumes were sliced so that the damage in the most critical areas, namely the tops and the bottoms of the bump/ball, could be calculated. The stress that initiates fatigue cracks occurs in these critical regions. The average damage in these portions of the solder correlates much better to the cycles to failure than the average damage over the entire bump/ball volume. Critical regions in the model were defined by 5 µm slices on the top and bottom of each bump while solder balls included 25 µm slices. Each bump/ball was attached to a copper pad at the top and bottom, which induces the CTE mismatch. The intermetallic layer (IMC) was not modeled, as it has shown in other studies to not significantly affect reliability predictions. *Figures 7 and 8*

show the detailed geometry of the bumps and balls used in this study.

Design of Experiments

Since the simulations examined three variables (lid CTE, PCB CTE, and the system ΔT) over a high/low range, a simple 2-level 3-factor DOE was used to create the simulation matrix. This required 8 simulations runs (2^3) along with an additional finer mesh simulation to validate mesh resolution. A full DOE identifies statistically significant variables and interactions, as well as the relative importance of variables. The test matrix produced outputs for the cycles to failure for both the solder bumps and solder balls. Those results are evaluated independently, so the identified significant variables and interactions may not be the same.

Table 1 shows the values of the parameters used in the DOE matrix. Values represent typical BGA manufacturing operating conditions, but with different material choices or operating environments, they easily could be larger/farther apart. Larger variable ranges would be better covered with a 3-level DOE.

The environmental cycle condition of 20°C or 50°C was applied to the entire BGA assembly. The die is not heated in this case (although it can be for actual problems). One thermal cycle was defined as a two-hour period with 15-minute dwell times at each extreme, and 45 minutes to traverse from one temperature to the other.

Run #	PCB CTE, ppm/K	Lid CTE, ppm/K	ΔT, °C
1	13	17	50
2	13	10	20
3	9.5	10	20
4	13	17	20
5	9.5	17	20
6	9.5	10	20
7	13	10	50
8	9.5	17	50

Table 1: DOE factors and runs

Conclusion

This article is part of a two-part series of articles based on Ref. 11. Part 2 of the series will describe the simulation software used, the mathematical model used for the solder, the analysis details, the statistical analysis of the DOE results, and conclusions drawn from the study, along with other details used in Part 1 of this article.

References

[1] Lee, W. W., Nguyen, L. T., & Selvaduray, G. S. (2000). Solder joint fatigue models: review and applicability to chip scale packages. Microelectronics reliability, 40(2), 231-244.

[2] Chen, G., Zhao, X., & Wu, H. (2017). A critical review of constitutive models for solders in electronic packaging. Advances in Mechanical Engineering, 9(8), 1687814017714976.

[3] Mukherjee, S., Nuhi, M., Dasgupta, A., & Modarres, M. (2016). Creep constitutive models suitable for solder alloys in electronic assemblies. Journal of electronic packaging, 138(3), 030801.

[4] Depiver, J. A., Mallik, S., & Amalu, E. H. (2020, September). Comparing and benchmarking fatigue behaviours of various SAC solders under thermo-mechanical loading. In 2020 IEEE 8th Electronics System-Integration Technology Conference (ESTC) (pp. 1-11). IEEE.

[5] Su, S., Akkara, F. J., Thaper, R., Alkhazali, A., Hamasha, M., & Hamasha, S. D. (2019). A state-of-the-art review of fatigue life prediction models for solder joint. Journal of Electronic Packaging, 141(4), 040802.

[6] Gabriel, O. E., and Huitink, D. R. (October 22, 2022). "Failure Mechanisms Driven Reliability Models for Power Electronics: A Review." ASME. J. Electron. Packag. June 2023; 145(2): 020801.

[7] Basit, M., Motalab, M., Suhling, J. C., & Lall, P. (2015, July). Viscoplastic constitutive model for Lead-free solder including effects of silver content, solidification profile, and severe aging. In International Electronic Packaging Technical Conference and Exhibition (Vol. 56895, p. V002T01A002). American Society of Mechanical Engineers.

[8] Syed, A., Sharon, G., & Darveaux, R. (2012, May). Factors affecting Pb-free flip chip bump reliability modeling for life prediction. In 2012 IEEE 62nd Electronic Components and Technology Conference (pp. 1715-1725). IEEE.

[9] Lau, John H., (April 13, 2022). Recent Advances and Trends in Advanced Packaging. Presented at IEEE EPS Binghamton Chapter (slide 11)

[10] Syed, A. (2004, June). Accumulated creep strain and energy density based thermal fatigue life prediction models for SnAgCu solder joints. In 2004 Proceedings. 54th electronic components and technology conference (IEEE Cat. No. 04CH37546) (Vol. 1, pp. 737-746). IEEE.

[11] J. Petroski and A. Carrasco, "Simulation of Solder Fatigue Effects on Typical BGA Package due to Material and Temperature Variations," *2024 40th Semiconductor Thermal Measurement, Modeling & Management Symposium (SEMI-THERM)*, San Jose, CA, USA, 2024, pp. 36-45.

[12] There are many dozens of papers with models considered, and each problem involves choosing the model best suited to the problem. One example of a good reference in this area is: Depiver, Joshua A., et al. "Creep damage of BGA solder interconnects subjected to thermal cycling and isothermal ageing." *2019 IEEE 21st Electronics Packaging Technology Conference (EPTC)*. IEEE, 2019.

Mitigating Flow Maldistribution in Data Center Two-Phase Cooling Systems with Flow Restrictors

Serdar Ozguc, Qingyang Wang, Akshith Narayanan, Richard W. Bonner III
Accelsius

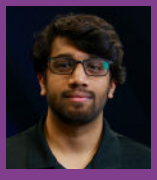
The Need for Two-Phase Cooling in Data Centers

Power usage in data centers account for 1.8% of the overall electricity expenditure in the United States [1] and the cooling infrastructures make up 50% of the total energy consumption of the data centers [2]. Power consumption translates to high operational costs and carbon footprint which can be potentially reduced by implementing higher efficiency thermal management systems. Moreover, the growing power densities in data centers, needed to address the demand for high-performance computing, are beginning to push the thermal limits of conventional air-cooled systems.

Liquid-cooled solutions were shown to significantly reduce the overall power consumption relative to air-cooling [3] and can handle increased power densities due to liquid's higher thermal conductivity and heat capacity. Single-phase liquid-cooling with

Serdar Ozguc
Dr. Serdar Ozguc earned his Ph.D. in Mechanical Engineering from Purdue University, where his research focused on the design and optimization of single- and two-phase microchannel heat sinks. He is currently a Senior R&D Thermal Engineer at Accelsius, where he specializes in the development of advanced two-phase direct-to-chip liquid cooling CDUs for data center applications.

Qingyang Wang
Dr. Qingyang Wang is a Senior R&D Thermal Engineer at Accelsius, where he is developing two-phase direct-to-chip cooling technologies for data centers. Prior to this role, he was an Associate Professor at North China Electric Power University. Dr. Wang obtained his Ph.D. in Mechanical Engineering from the University of California, San Diego, in 2020, and his B.S. in Thermal Engineering from Tsinghua University in 2015. With over 40 peer-reviewed publications in leading journals and conferences, he is recognized for his expertise in multiphase flow and heat transfer, thermal management, and thermal characterization.



Akshith Narayanan
Akshith Narayanan is a recent Masters Graduate in mechanical engineering from Georgia Institute of Technology. He completed his degree with a thesis working on an investigation of near-junction flow boiling of high-power electric vehicle inverters. He joined Accelsius, an emerging two phase direct to chip liquid cooling company looking to provide an elegant two-phase solution for data center high power chips. He has years of experience in heat transfer, fluid mechanics with specific expertise in two phase boiling.

Richard W. Bonner III
Dr. Bonner is a distinguished heat transfer researcher with 20 years in thermal product development, specializing in two-phase cooling. He has authored over fifty papers and holds five U.S. patents, having designed cooling products for 125+ clients in various sectors. He is a former AIChE Transport and Energy Processes Division Director and holds a B.S., M.S., and Ph.D. in Chemical Engineering from Lehigh University.

water-based coolants can achieve good thermal performance due to water's favorable thermal characteristics. However, a minor leak in a water-cooled system can cause a catastrophic electrical failure, due to water's unfavorable electrical characteristics. Two-phase liquid-cooling with refrigerants circumvents this problem where high heat transfer performance is achieved through liquid-to-vapor phase change of the dielectric fluid. Heat fluxes on the order of 1 kW/cm^2 have been dissipated using two-phase cooling [4].

Flow Maldistribution in Two-Phase Cooling

Pumped two-phase cooling has been studied extensively in the literature [5]. However, implementation and testing of two-phase cooling in data center applications are limited. The highly parallel architecture of liquid-cooling loops in server racks can suffer from coolant flow maldistribution between heat generating components. This problem is more prominent in two-phase flows because the difference of pressure drop between liquid and vapor flows can lead to instability. High heat loads on cold plates result in increased vapor generation, leading to a rise in pressure drop. This, in turn, diverts the coolant through the cold plates with lower heat loads, potentially causing device overheating.

Flow restrictors can be adopted upstream of boiling [6] to suppress maldistribution by increasing the liquid line pressure drop relative to the vapor line. However, careful design of restrictors is crucial to effectively mitigate maldistribution without inducing excessive pressure drops that reduce the overall flow rate of the system.

Rack-Level Two-Phase Cooling System

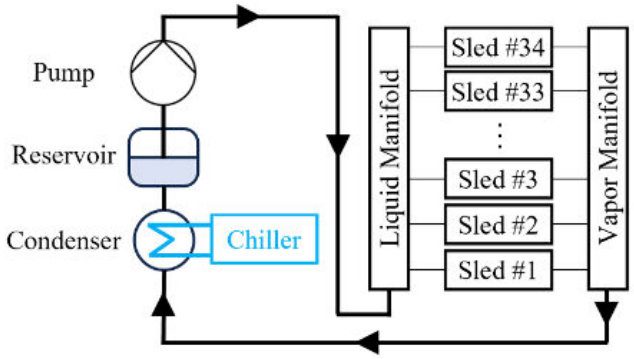


Figure 1: Flow diagram of the investigated server-rack level two-phase flow loop for data center cooling

Pumped two-phase flow loop for a data center rack consisting of 34 server sleds with heat generating components was investigated. A flow diagram of the loop is shown in *Figure 1*. The heat dissipation from each sled is between 0-2 kW and can vary among sleds. Refrigerant R-1233zd(E) was used as the two-phase coolant in the main loop. Pumped coolant entered the sleds as a liquid, absorbed the generated heat, and left as a liquid/vapor mixture with thermodynamic qualities of 0-85%. The sleds were connected to the loop via a liquid manifold upstream and a vapor manifold downstream. The liquid and vapor manifolds are large pipes (1" and 2" inner diameter respectively) connected to the sleds with additional tubing.

Hydrodynamic Maldistribution Model

A numerical model of the liquid manifold, vapor manifold, and the server sleds was developed. In this model, two-phase flow is simplified using the homogenous flow assumption, wherein the vapor bubbles and the surrounding liquid move at the same velocity. The liquid and vapor manifolds are discretized along the length, and the conservation equations for mass, momentum, and energy are applied. The frictional losses in the manifolds are estimated using friction factor correlations for fully developed laminar [7] and turbulent flows [8] in circular channels alongside with a mixture viscosity correlation [9].

Pressure drop between the liquid and vapor manifolds ($\Delta P_{manifold}$) at a discretized position j along the length depends on the pressure drop of the sled components in between (ΔP_{sled}), which include hoses, tubes, fittings, cold plates, etc. and the pressure drop across the flow restrictors ($\Delta P_{restrictor}$) placed upstream of the cold plates as shown in *equation 1*.

$$\Delta P_{manifold} = \Delta P_{sled} + \Delta P_{restrictor} \quad (1)$$

Sled Pressure Drop and Thermal Resistance

Two-phase pressure drop of an individual sled was experimentally measured to derive an empirical correlation. A flow diagram of the experimental flow loop is shown in *Figure 2*. The investigated server sled consists of two heat generating components that were emulated using two thermal test vehicles (TTV) with heat spreaders and cold plates. Pressure drop and temperature data were collected over ranges of flow rates (5-26 g/s) and total heat inputs (0.2-2 kW), resulting in thermodynamic qualities between 0 and 100%[1]. A total of 104 data points are collected.

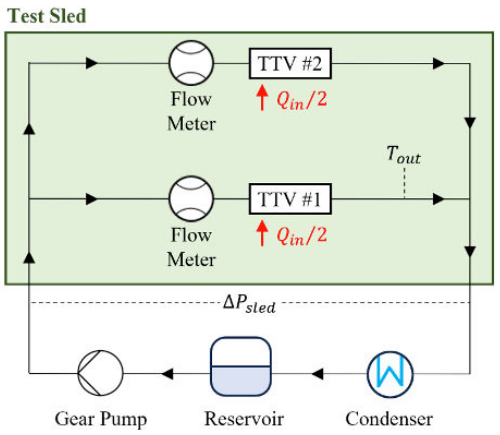


Figure 2: Flow diagram of the experimental flow loop for testing pressure drop across a server-sled

[1] Since the thermodynamic quality is difference between fluid and saturated liquid enthalpies divided by the heat of vaporization: $x = (h - h_f)/h_{fg}$, with superheating, x can be >100%.

A multi-variable, second-order polynomial fit was used to generate an empirical correlation for the pressure drop data. The resulting correlation for sled pressure drop (ΔP_{sled}) matches quite well with the measured data, with most of the predictions having less than 25% error. *Figure 3* shows thermal resistance (normalized by the lowest measured resistance) at exit thermodynamic qualities between 0-100% for all flow rates. The thermal resistance is lowest at 54% thermodynamic quality and sharply increases near 0% and 100%, as expected. Thermal performance improves from single-phase liquid cooling, with near zero thermodynamic qualities, to higher thermodynamic qualities as flow boiling enhances convection. However, thermal resistance increases near x_{sled} = 100% because the generated vapor starts interfering with heat transfer. The two-phase loop shown in *Figure 1* was designed to operate at an exit thermodynamic quality of 70% to provide a margin of safety for reliability. However, individual sleds are expected to have higher qualities due to flow maldistribution. An upper limit of x_{target} = 85% was chosen to ensure acceptable cooling performance in the sleds.

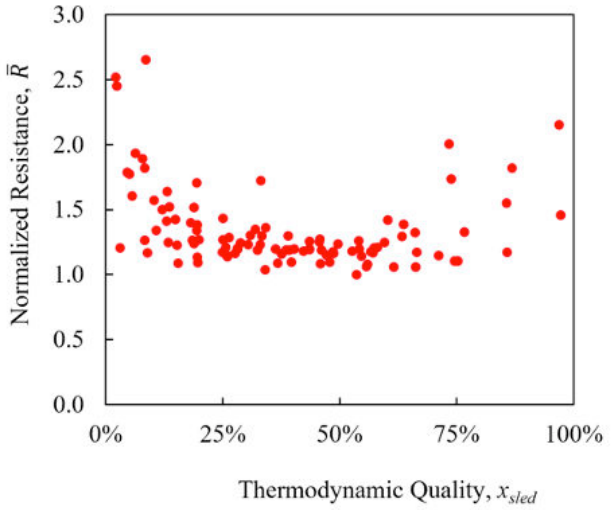


Figure 3: Experimental thermal resistance vs. exit thermodynamic quality for all flow rates tested

Restrictor Pressure Drop

Two restrictors were placed in each sled, positioned upstream of the cold plates. *Equation 2* was used to model the normalized pressure drop across restrictors ($\Delta\bar{P}_{restrictor}$) with respect to the normalized mass flow rate ($\bar{m}$).

$$\Delta\bar{P}_{restrictor} = \alpha\bar{m}^{\beta} \quad (2)$$

Term α is the flow resistance factor and β is the flow scaling exponent. Pressure drop and mass flow rate are normalized so that the flow resistance factor (α) is the ratio of restrictor pressure drop to sled pressure drop under maximum heat load (2 kW). Values of α and β are dictated by the geometry of the restrictor and coolant properties:

- Linear scaling (β = 1) occurs in a long tube with a small diameter where viscous shear dominates the pressure drop.
- For orifice restrictors, pressure drop scales quadratically with mass flow rate (β = 2) due to a momentum dominated flow, and the value of α depends on the orifice size.
- Higher order scaling can be achieved through moving or flexible parts in the restrictors. For example, commercial flow regulators incorporate flexible polymers that constrict the flow area with increasing pressure differential.

Nonuniform Heating

The sleds might be under different loads during operation. Therefore, the two-phase flow was evaluated using the hypothetical heating profile of *equation 3*.

$$Q_{in} = 2\ kW \cdot \left(\frac{y}{H}\right)^2 \quad (3)$$

In this equation, y is the distance from the bottom sled and H is the total height between 34 sleds, which are stacked on top of each other. The sled at the bottom receives no heat while the sled at the top receives 2 kW.

Flow Maldistribution Without Restrictors

The two-phase flow loop was analyzed without flow restrictors to serve as a benchmark. The inlet flow rate to the liquid manifold was selected such that each sled should have an exit quality of 70% in the absence of maldistribution. *Figure 4* shows the resulting exit thermodynamic qualities for each sled, with higher exit thermodynamic quality in the sleds near the top. In addition to the gravitational effects, the flow maldistribution is exacerbated by the nonuniform heating. Since pressure drop increases with vapor generation, the sleds with higher heat input near the top receive less flow. The maximum predicted thermodynamic quality is 176%, which indicates that the cold plate is under dry-out that leads to a high thermal resistance that is unacceptable for data center cooling.

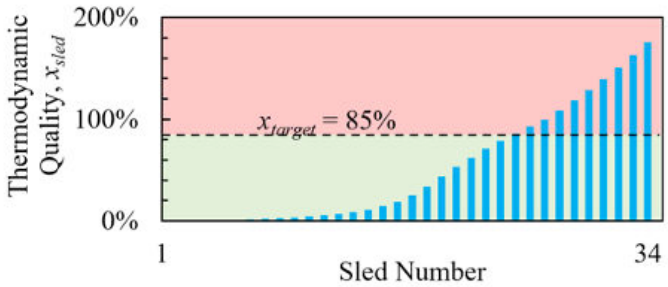


Figure 4: Predicted exit thermodynamic quality distribution under nonuniform heating (*Equation 3*) without flow restrictors

Flow Maldistribution with Orifice Restrictors: An orifice restrictor (β = 2) was first investigated to suppress maldistribution. An insufficiently low α value cannot suppress the maldistribution

while an unnecessarily high value will induce a high pressure drop penalty. *Figure 5* shows the thermodynamic quality distribution for $\alpha = 2.0$. The maximum thermodynamic quality is 85% under nonuniform heating. Therefore, all the sleds operate under the maximum thermodynamic quality limit. Flow resistance factor (α) is the ratio of restrictor to sled pressure drop. For the orifice restrictor, the pressure drop across the orifice needs to be roughly twice as much as the pressure drop across the server sled components.

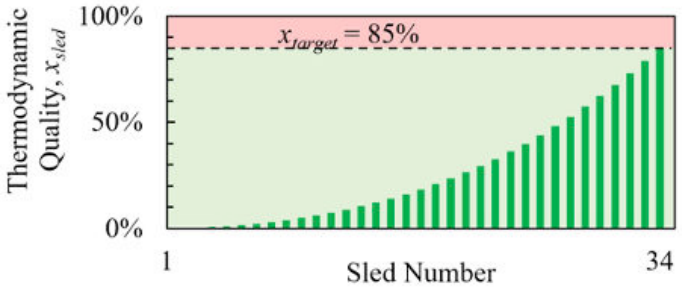


Figure 5: Predicted exit thermodynamic quality distribution under nonuniform heating (*Equation 3*) with orifice restrictors

Advantage of Higher Order Restrictors

The two-phase flow loop was analyzed for a range of β values. For each case, the value of α required to achieve a maximum 85% thermodynamic quality under nonuniform heating was found. Resulting α and β pairs are shown in *Figure 6*. The value of α, and hence the restrictor pressure drop, decreases with increasing β. The increasingly concave pressure drop-mass flow rate response severely punishes maldistribution, pressure drop is reduced at and below the desired flow rate. The value of α converges to 0 as $\beta \rightarrow \infty$. An ideal flow regulator can suppress the maldistribution without inducing additional pressure drop to the system at the desired flow rate and maximum heat input.

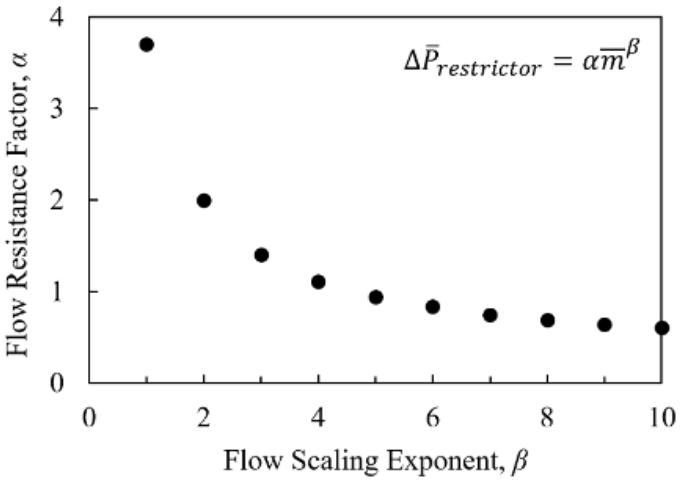


Figure 6: Flow resistance factors to obtain a maximum vapor exit quality of 85% at various flow scaling exponent values

Conclusions

Lower pressure drop restrictors in a pumped two-phase loop can enable a higher flow rate across the system to dissipate more heat. Therefore, restrictors with higher scaling exponents (β), such as flow regulators, are preferable. However, there are practical challenges that make their implementation difficult. First, commercially available flow regulators are significantly more costly compared to off-the-shelf orifice restrictors. Second, commonly adopted orifice restrictors suppress two-phase backflow by inducing high pressure drop upstream, which is not addressed by flow regulators Third, flow regulators need to be highly tailored for the system, which requires a good understanding of the entire pressure drop response of the flow loop under different conditions. If the available flow rate is underestimated, a flow regulator would completely block the additional flow, thereby compromising the added cooling capacity. If the available flow is overestimated, a flow regulator would not effectively suppress maldistribution. An orifice restrictor is more robust to uncertainty, providing reliable suppression to maldistribution when the system characteristics are not fully characterized.

References

[1] M. A. B. Siddik, A. Shehabi, and L. Marston, "The environmental footprint of data centers in the United States," *Environ. Res. Lett.*, vol. 16, no. 6, p. 064017, Jun. 2021, doi: 10.1088/1748-9326/abfba1.

[2] M. Dayarathna, Y. Wen, and R. Fan, "Data Center Energy Consumption Modeling: A Survey," *IEEE Commun. Surv. Tutor.*, vol. 18, no. 1, pp. 732–794, 2016, doi: 10.1109/COMST.2015.2481183.

[3] F. Rebarber, "Quantifying the Impact on PUE and Energy Consumption When Introducing Liquid Cooling Into an Air-cooled Data Center." Accessed: Jan. 01, 2024. [Online]. Available: https://www.vertiv.com/en-emea/about/news-and-insights/articles/blog-posts/quantifying-data-center-pue-when-introducing-liquid-cooling/#:~:text=Impact%20of%20the%20Introduction%20of,compared%20to%20100%25%20air%20cooling

[4] R. K. Mandel, D. G. Bae, and M. M. Ohadi, "Embedded Two-Phase Cooling of High Flux Electronics Via Press-Fit and Bonded FEEDS Coolers," *J. Electron. Packag.*, vol. 140, no. 3, p. 031003, Sep. 2018, doi: 10.1115/1.4039264.

[5] 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.
[6] W. Chen, T. M. Conboy, G. W. Daines, and D. W. Fogg, "A Robust Two-Phase Pumped Loop with Multiple Evaporators and Multiple Radiators for Spacecraft Applications," presented at the 47th International Conference on Environmental Systems, Charleston, South Carolina, Jul. 2017.
[7] *F. P. Incropera, D. P. DeWitt, T. L. Bergman, A. S. Lavine, "Fundamentals of Heat and Mass Transfer," Wiley, 2011.*
[8] V. Mileikovskyi and T. Tkachenko, "Precise Explicit Approximations of the Colebrook-White Equation for Engineering Systems," in *Proceedings of EcoComfort 2020*, vol. 100, Z. Blikharskyy, Ed., in Lecture Notes in Civil Engineering, vol. 100. , Cham: Springer International Publishing, 2021, pp. 303–310. doi: 10.1007/978-3-030-57340-9_37.
[9] "A, Cicchitti, C. Lombardi, M. Silvestri, G. Soldaini, R. Zavattarelli, 'Two-phase cooling experiments – pressure drop, heat transfer and burnout measurements,' Energia Nucleare, 7 (1960) 407-425.".

Augmenting Development of Electronics Cooling Technologies with Machine Learning Tools – a Heat Pipe System Example

Van P. Carey
Mechanical Engineering Department
University of California at Berkeley

The traditional pathway for development of thermal management technologies for electronics has been to analyze using heat transfer theory augmented by computational tools such as finite difference or finite element methods, CFD tools, or thermal management system simulation tools, sometimes in tandem with prototype system fabrication and testing. Combining recently available high speed computing processors, faster memory, and machine learning strategies now offers ways to enhance these traditional tools for thermal control component and system development. These trends are producing an evolution of thermal management technology development towards use of physics-inspired machine learning approaches that combine machine-learning with heat transfer theory.

To illustrate this point, the development of a heat pipe system for a server machine is considered. The example system of interest, shown schematically in *Figure 1*, has two evaporators and a single condenser.

This is a simplified example of a real heat pipe system that could have multiple evaporators with a single condenser heat rejection heat exchanger. Note that in this example heat pipe system, operating conditions correspond to the specified values of condenser inlet water temperature T_{cfi} and flow rate $\dot{m}_c$, and the heat rejection rates in the evaporators $\dot{q}_a$ and $\dot{q}_b$. If we combine the governing heat transfer relations and conservation equations indicated in *Figure 1*, the performance parameters, the temperatures of computer chips *a* and *b* ($T_{chip,a}$ and $T_{chip,b}$) can be computed if the conductances ($UA_{e,a}$, $UA_{e,b}$, UA_c) are known.

Modeling performance of this system can be accomplished with physical modeling alone if the conductances can be predicted with submodels, or determined from experiments. These conductances are typically the result of combined conduction and convection effects, and the geometry of the components may require multidimensional analysis. Multiphysics modeling tools (ANSYS, COMSOL, etc.) could be used to determine the conductances, or separate experiments on the device components could be used to determine conductance values. These approaches are well known and extensively used in traditional thermal control system development.

The availability of machine learning approaches opens the door to at least two other approaches: (1) using physical modeling with a genetic algorithm and (2) modeling using a neural network data-based model. The features and pros and cons of each of these are discussed below.

A Physics Inspired Model and Genetic Algorithm

In this model it is presumed that there is a data set in which each point is a list of variables that includes the operating conditions

Van P. Carey
Professor Carey is widely recognized for his research on near-interface micro- and nanoscale thermophysics and transport in liquid-vapor systems, and computational modeling and simulation of energy conversion and transport processes. His research has frequently included modeling at multiple scales, ranging from the molecular level (molecular dynamics simulation of thermophysics) to the device and system level (multidevice system models). His research is also exploring the use of machine learning strategies to enhance performance of energy conversion and transport in applications, and create energy technologies that can autonomously adapt to maximize their performance and reduce their environmental impact.

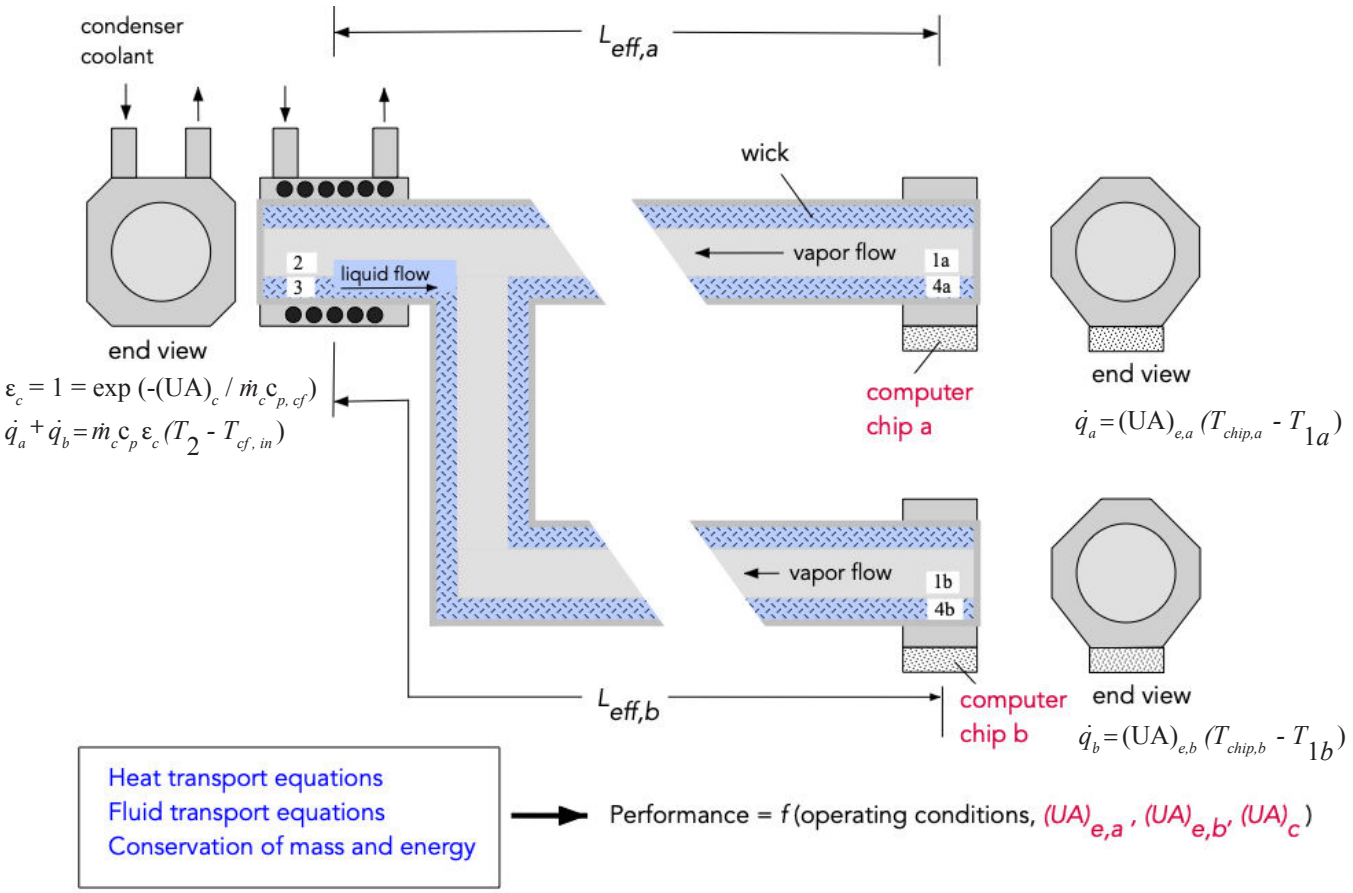


Figure 1: Heat pipe example system model

variables X_i and the resulting performance parameter values Y_j for those conditions: [X_1, X_2, X_3, … X_m, Y_1, Y_2, Y_3, … Y_n].

Here the recommended practice of normalizing the data would be applied, with each variable being normalized with the median value for that parameter in the data set before it is input to the model, and the model is trained to predict the normalized chips temperatures, which equal the respective actual chip temperature divided by their median value in the data set. Once the model is trained in this way, to predict performance, the operating conditions are normalized with the corresponding median value used in training, and the model predicts the normalized chip temperatures. The physical chip temperatures are then determined by multiply each by their corresponding median temperature used in training.

The model framework discussed above dictates that once a set of conductances are specified, the model can predict the performance parameters for a given set of input parameters. This perspective is the basis for using a genetic algorithm model. In such a strategy, we want to find the set of set of ($UA_{e,a}$, $UA_{e,b}$, UA_c) that best fit the data set. The genetic algorithm accomplishes this in the fmanner describe below.

Genetic Algorithm Structure

An initial population (ensemble) of solutions (sets of ($UA_{e,a}$, $UA_{e,b}$, UA_c) combinations here) is established. Each individual is a candidate solution to the problem (a set of ($UA_{e,a}$, $UA_{e,b}$, UA_c) values) analogous to a biological organism in a population of a specie, which is characterized by a set of parameters (variables) known as genes. Genes are joined into a string to form a chromosome (solution). Once the population is established, the following steps are iteratively executed:

(i) Fitness function determination

To begin each iteration, the algorithm computes a fitness score for each individual. Here that will be done by randomly pairing the solution with one point of the database, computing output Y_i variables for that point's input variables, and computing the RMS fractional error between the solution predicted Y_i values and the values in that data point. The probability that an individual will be selected for reproduction is based on its fitness score – the lower the RMS fractional error, the higher the survival probability.

(ii) Selection of fittest to survive each generation

Individuals with a fitness score below a threshold are eliminated from the population and replaced with offspring.

(iii) Survivor offspring production with gene choice from surviving parents and mutation

In new offspring formed, some of their genes can be subjected to a random mutation with a low random probability.

Steps (i)-(iii) are repeated for successive generations until the RMS fractional error (fitness function) for population reaches a sufficiently low value.

The details of how the above steps are handled may vary somewhat in genetic algorithm applications, but the elements of the genetic algorithm generally conform to the features described above. In the example calculation summarized here, as described above, a data set of normalized performance data points were used, each having the form: [X_1, X_2, X_3, … X_m, Y_1, Y_2, Y_3, … Y_n], where X_i are the operating conditions variables and Y_j are the resulting performance parameter values.

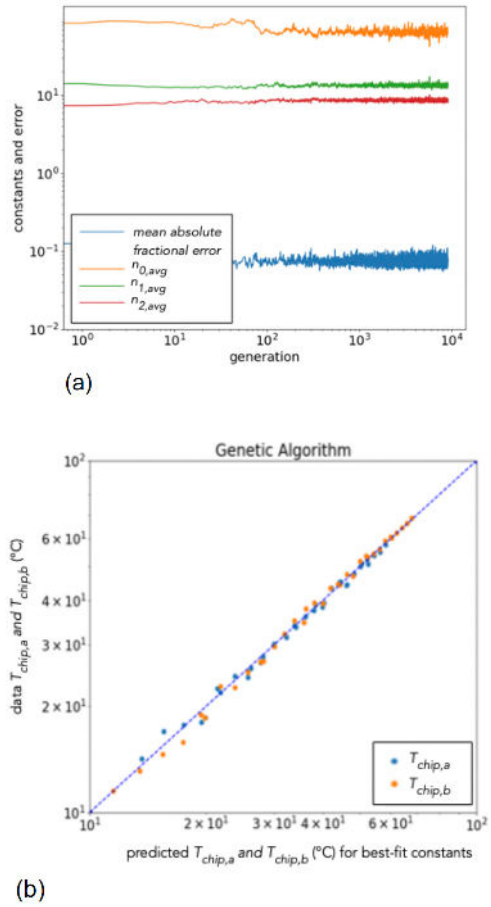


Figure 2

A genetic algorithm with the features described above was used to determine the set of conductances ($UA_{e,a}$, $UA_{e,b}$, UA_c) that minimized the mean RMS error between predictions of the model and randomly selected data points in the population of gene sets ($UA_{e,a}$, $UA_{e,b}$, UA_c). *Figure 2a* shows computational convergence of the mean gene values as successive generations are analyzed. The resulting best fit constants were used in the model equations to predict the two chip temperatures, and comparison of the predictions with data chip temperatures for the same operating conditions is shown in *Figure 2b*. The resulting model with the fitted constants agrees with the data to a mean absolute fractional error of about 0.075.

The results in *Figure 2b* are for the following genetic algorithm model determined mean (generationaveraged) values providing a best fit: $(UA)_{e,a}$ = 65.32 W/K, $(UA)_{e,b}$ = 13.44 W/K, $(UA)_c$ = 8.66 W/K. These values resulted in a fit with a mean fractional error of 0.075 (~ 7.5%).

Performance Prediction with a Conventional Neural Network Data Based Model

A conventional neural network can be trained to predict the trends in output performance parameters for prescribed input parameters using performance data like that described in the genetic algorithm model described above. In the simple example system of interest here, the input parameters are the operating conditions corresponding to the specified values of condenser inlet water temperature T_{cfi} and the heat rejection rates in the evaporators $\dot{q}_a$ and $\dot{q}_b$. The output performance parameters are the chip operating temperatures: $T_{chip,a}$ and $T_{chip,b}$. The neural network model is shown schematically in *Figure 3*.

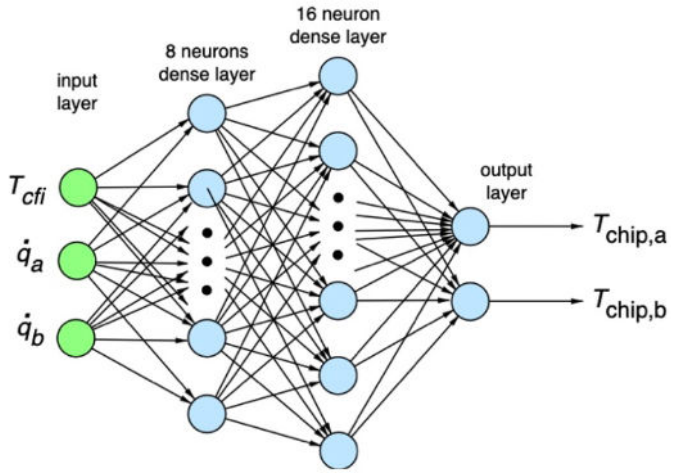


Figure 3

Note that the neural network has three inputs, one output layer, and two layers between. The presence of these "hidden" layers is sufficient to categorize this as a deep learning model. Each neuron in the model multiplies each input it receives by an adjustable constant, sums them, and adds a biasing constant which is adjustable. The result is handed to a specified activation function. Here that is chosen to be a rectilinear exponential linear unit (RELU) function [1]. Although the network structure for this example is simple, this model has 210 adjustable parameters, which allows it to model fairly complex, non-linear, multivariate behavior in the data set. The code to set up the model and train it was set up using keras [2] and other standard python tools. This model was

trained with the same data used in the genetic algorithm model described above. The trained model was used to predict the two chip temperatures, and comparison of the resulting predictions with data chip temperatures for the same operating conditions is shown in *Figure 4*. The neural network model fits the data to a mean absolute fractional error of 0.018 (~1.8%).

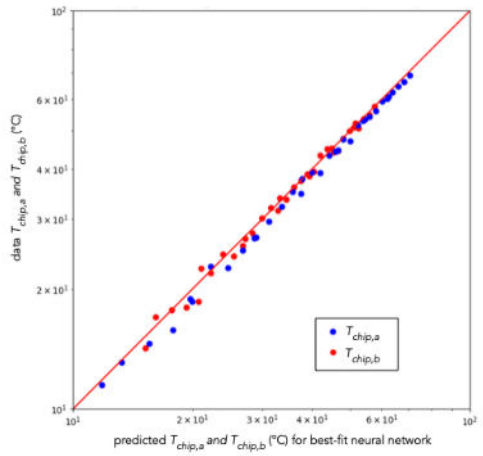


Figure 4

Once the model is trained, its learned knowledge of the parametric performance trends in the data is stored in its 210 neuron parameter values. It can then be used to predict performance for an arbitrarily chose set of operating conditions. This is illustrated in *Figure 5*. For a heat pipe condenser coolant temperature of 10°C, this figure shows the model predicted variations of the two chip temperatures as functions of the heat power generation in each of the two chips A and B. Note that for each pair of heat dissipation rates experienced by the system, the corresponding points on these surfaces predict the operating temperature of each of the chips.

Closure: Pros and Cons

The results of the example models presented here demonstrate that either of the two machine learning methodologies described here could be a useful methodology for predicting performance of the heat pipe system shown in *Figure 1*. Both of these are data-based models, so they both would require fabrication and testing of a prototype heat pipe system to create the database to train the model. Here are some other pro's and con's of each:

Physics-Inspired Genetic Algorithm

Once trained, the genetic algorithm physics-inspired model is embodied in a set of mathematical equations with known coefficients that can be used to predict the heat pipe system performance. This model has the further advantage that it back-infers parameters in the model that might be difficult to predict from theory, difficult to measure separately, and/or involve physical properties that are not accurately known. Knowledge of the inferred conductances may be very valuable in that they may indicate that one or more conductances are much different than expected from heat transfer theory, implying that a change in the design, material, or manufacturing process is needed to achieve the desired level of performance. Also, if operating conditions and performance are monitored during field operation, this heat pipe model can be retrained to reflect shifts in performance due to condenser water side fouling or shifts in contact resistance at the chip-evaporator interface. This can facilitate model-based control of the condenser cooling water flow rate and inlet temperature for appropriate chip cooling. A drawback of this type of model is that it assumes that the conductance parameters are fixed constants for all of

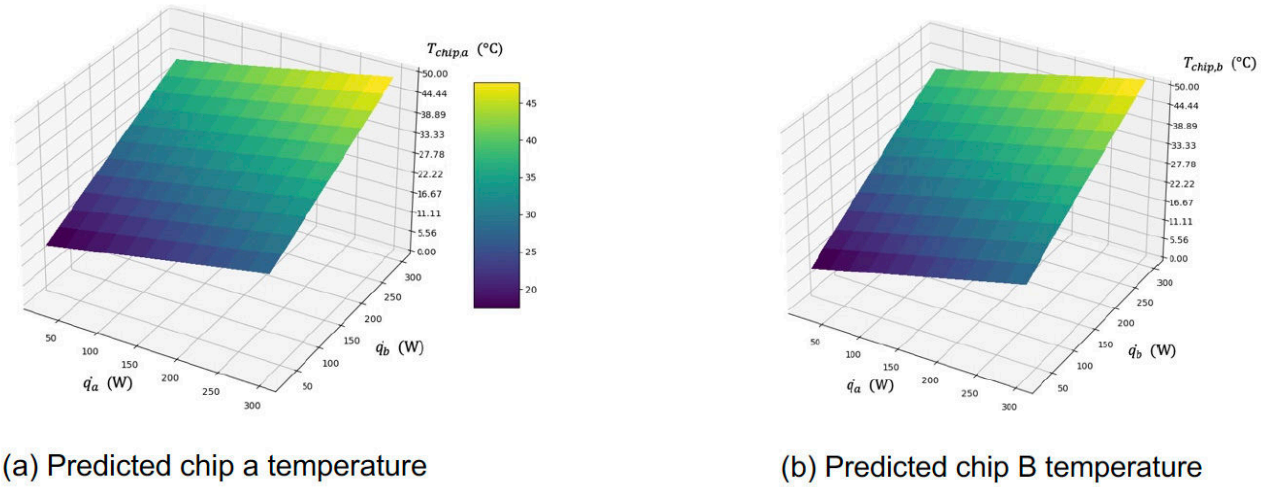


Figure 5: Predicted variations of chip temperatures with chip heat dissipation for a condenser coolant temperature of 10°C

the operating conditions of interest, which may not be fully accurate. Adjustments the model to account for parameter variation wit operating conditions could be made.

Artificial Neural Network Model

Once trained, the neural network model can be used to predict performance for the heat pipe system, but to do so, the neural network model code must be used with the best-fit values of neuron weights and biases determined in the training. This model can fit trends in the data with a higher degree of adaptability than the specific model relations used in the genetic algorithm model. However, this type of model does not provide an explicit mathematical relation for the performance as a function of operating condition. Consequently, to define the parametric trends for the system, the model must be used to generate predictions over the parameter space of interest and the results must be analyzed to assess the trends.

While only two types of machine learning tools have been explored here, the example results and observations reflect the potential usefulness and advantages of using tools of this type to enhance development of better-performing systems, and/or facilitating adaptive control of heat pipe based thermal control systems for electronics applications.

References

[1] https://keras.io/api/layers/activations/

[2] https://keras.io/guides/

Hydraulic and Thermal Characteristics of a Double-Sided Cold Plate – Part 2

Tesla Tear Down, CFD Validation, and Machine Learning to Determine the Performance Limit

Azita Soleymani
HeatSync

John Wilson
Siemens Digital Industries Software

William Maltz
Electronic Cooling Solutions, Inc.

This study investigates the hydraulic and thermal characteristics of the TESLA AUTOPILOT HW2.5 MODEL 3 Y, which features a double-sided cold plate with PCBs attached on both sides. We present a step-by-step teardown process of the unit, measuring internal dimensions, fin locations, inlet/exhaust ports, pedestals, and cold-plate channels. These measurements enabled the creation of a 3D numerical model for the Tesla AUTOPILOT HW2.5 unit, which was used to analyze the hydraulic and thermal performance of the cold plate.

We conducted experiments to measure hydraulic head loss through the cold plate and evaluate thermal performance at various coolant flow rates and power loads. These results were used to validate the CFD model. Thermal loads and the fluid flow rate

Azita Soleymani
Dr. Azita Soleymani is the CTO of HeatSync, empowering excellence through thermal innovation in electronics and battery systems. With nearly 150 years of combined expertise, HeatSync (https://heat-sync.com/) tackles some of the most demanding thermal challenges across industries, including consumer electronics, automotive, energy storage, telecommunications, aerospace, medical devices, data centers, AI systems, and more. Formerly a member of the technical staff at Meta, Dr. Soleymani has co-authored over 40 technical papers and is a recognized leader in thermal design, simulation, testing, and optimization.

John Wilson
Electronics Thermal Application Specialist at Siemens, Digital Industries Software. John Wilson joined Mentor Graphics Corporation, Mechanical Analysis Division (formerly Flomerics Ltd), that is now part of Siemens Digital Industries Software, in 1999 and has over 25 years of thermal design experience relating to simulation and testing. John currently works with product management teams to provide electronics thermal design solutions to leading electronics industry clients globally.

William Maltz
William Maltz has 38 years of experience. The ECS team, he leads, provides design advice to clients. ECS uses numerical analysis and experimental work to identify design solutions. He chaired technical sessions at SemiTherm, IMAPS, InterPACK, ITherm and co-authored technical papers and magazine articles on CFD, liquid cooling, natural convection, optimization techniques, and boundary condition independent models. He has reviewed books, publications, and presentations. William mentored 20+ engineers, who utilized their ECS experience to launch careers elsewhere. He is an ASME Fellow and active in the local ASME Section.

in the CFD model mimicked the experimental test cases that were evaluated. The model results for the hydraulic pressure loss and temperatures at specific locations were found to correlate well with the experimental data that was collected. The validated CFD model was then used to examine the thermal characteristics of the cold plate across a range of operating conditions. The model was also used to identify opportunities to improve the design of the cold plate.

Additionally, an Artificial Neural Network (ANN) model of the cold plate assembly was developed to analyze system performance across *a wide range* of operating parameters. It was demonstrated that the developed neural network model could be used to determine the performance limit of the cold plate.

Keywords: Cold plate, Autonomous driving, CFD, Tear down, Thermal performance, Hydraulic performance, Artificial Neural Network, Machine Learning.

Introduction

Part 1 of this paper [1] detailed the construction and simulation results of the CFD model of the cold plate assembly, providing insights into its performance through velocity and temperature profiles. Part 2 outlines the teardown process and experimental setup used to validate the simulation model.

Artificial intelligence (AI) and machine learning (ML) have gained prominence in both scientific and public discourse over the past decade. ML algorithms utilize data subsets to generate predictive rules for system outcomes based on input variables. The ANN, a subset of ML, learns from example data to form probability-weighted associations between inputs and results that are stored within its structure.

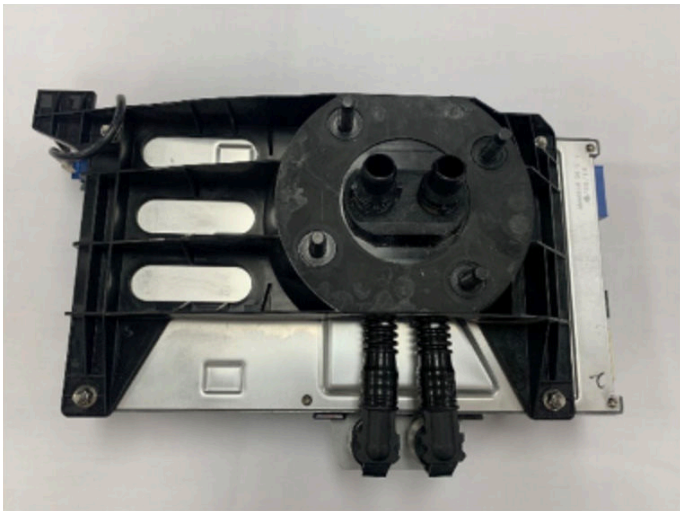
In this paper, an ANN of the cold plate assembly is presented to study the performance of the system using multiple combinations of the operating parameters.

Tesla Autopilot Assembly Tear Down

The autopilot unit, shown in *Figure 1*, is enclosed in an aluminum housing and with a cold plate transition assembly attached to the front side of the housing.

First, the autopilot cold plate transition assembly, which includes the coolant hoses shown in *Figure 2*, was detached from the housing of the autopilot assembly. The hoses are essential components of the cooling loop, allowing coolant to flow into and out of the cold plate. Additionally, *Figure 2* illustrates the attachment points for the hoses on the transition assembly, providing a clear view of the pathways for coolant circulation within the system.

Then, the top and bottom sections of the autopilot aluminum housing were removed. As shown in *Figure 3*, the cold plate serves as the central structure within the assembly, with printed circuit boards (PCBs) securely mounted on both sides.



Front

Back

Figure 1: The autopilot assembly before teardown

As shown in *Figure 4*, these PCBs house critical components, including four high-power GPUs or processors, which are strategically positioned for optimized thermal management. Additionally, various support electronics and connectors are visible on the PCBs, designed to facilitate signal processing and power management for the autopilot system.

The cold plate, after the thermal putty was removed, is shown in *Figure 5*. Finned channels are integrated into the cold plate design to improve the heat transfer. There are four lapped, or polished, pedestals with the four main chips attached to them. These pedestals are labeled A, B, C, and D to provide reference points for the discussion in this document.

The cold plate was dismantled so that the dimensions of the internal geometry could be measured. The measurements were used to generate a CAD model of the unit.

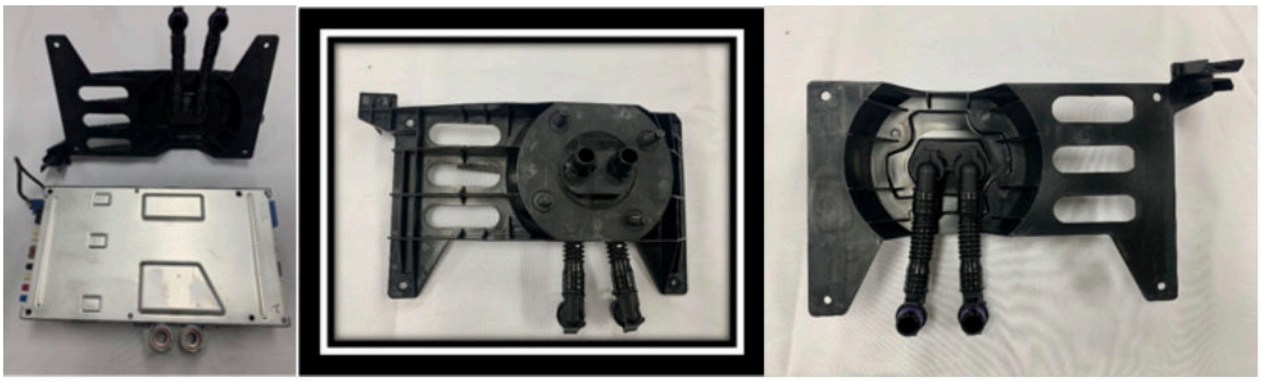


Figure 2: The hoses are attached to the autopilot cold plate transition assembly

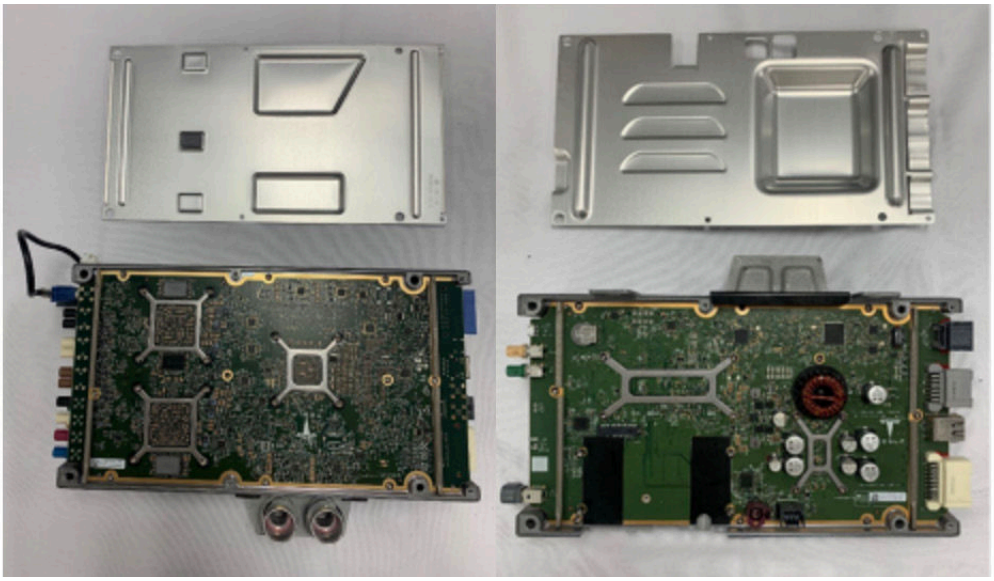


Figure 3: The autopilot consists of two PCBs attached to both sides of a cold plate

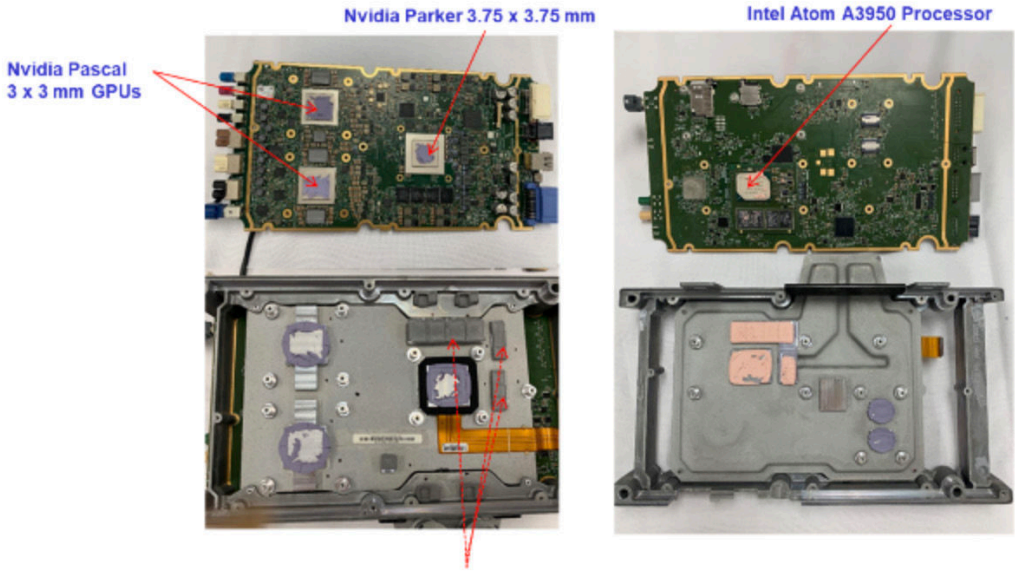


Figure 4: Removal of the PCBs from the Cold Plate

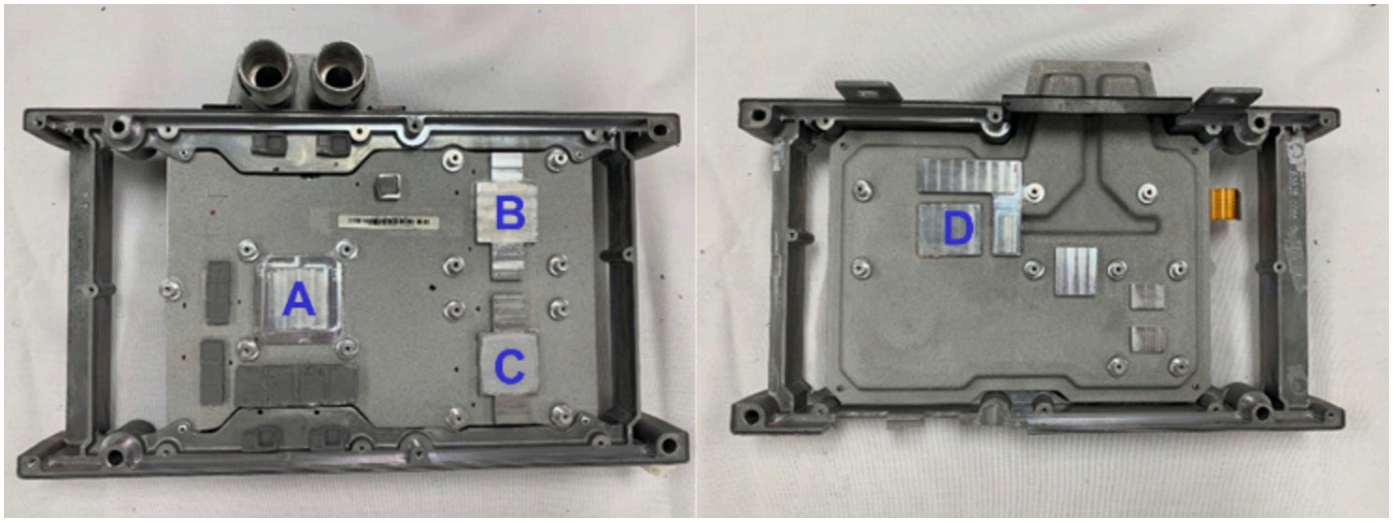


Figure 5: Front and Back Views of the Die Casted Cold Plate

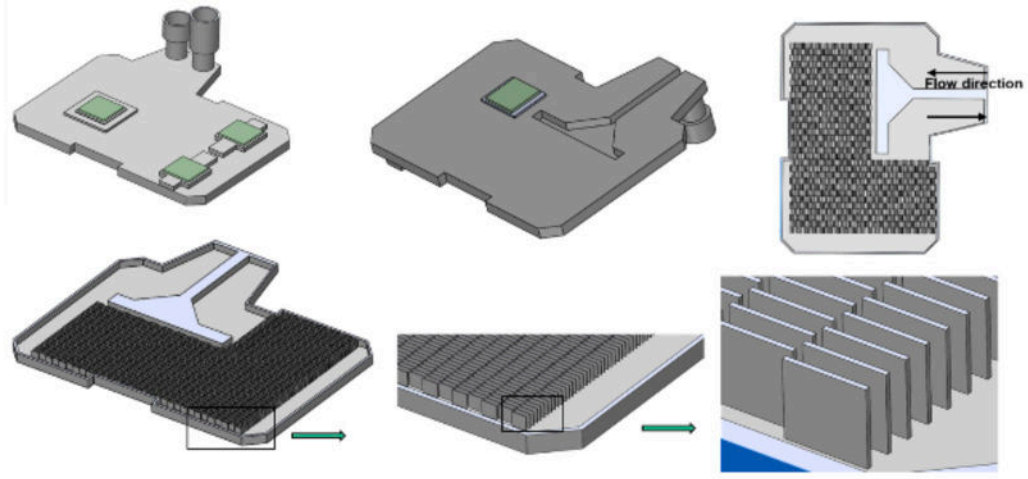


Figure 6: The CAD Model Representing the Double-sided Cold Plate

Experimental Setup

A set of experiments were conducted to measure the hydraulic head loss through the cold plate and assess its thermal performance under varying coolant flow rates and power loads. To facilitate these tests, a second autopilot unit was disassembled, and thermocouples and heaters representing the four main chips were attached to the cold plate. The coolant used in these experiments was a 50/50 glycol-water mixture, consistent with the specifications provided by the original design. This coolant choice aligns with the intended operating conditions, ensuring that test results accurately reflect the cold plate's performance in real-world applications.

Temperatures of the pedestal below each heater module pressure head loss through the cold plate, and the liquid temperatures were all measured. Uniform power was applied to the four heaters.

The heaters mounted to the cold plate were driven by a variable power supply so that a specified power dissipation could be applied to the cold plate. The four heaters were placed in the locations where the GPUs and processors contact the cold plate. *Figure 7* shows locations where thermocouples and heaters were attached to the front side of the cold plate.

Arctic Silver compound was placed on the interfaces between the heaters and the cold plate pedestals. A 2 mm x 1 mm groove was milled from the side to the center of each pedestal. The thermocouple beads were then secured in the groove with Arctic Alumina epoxy. After curing, the grooves were filled with additional Arctic Alumina epoxy, and a wood insulator was placed over each heater to improve thermal control. The entire assembly was then thermally insulated; however, while insulating the heaters themselves enhanced measurement accuracy, insulating the full assembly had a negligible effect on accuracy within the range studied.

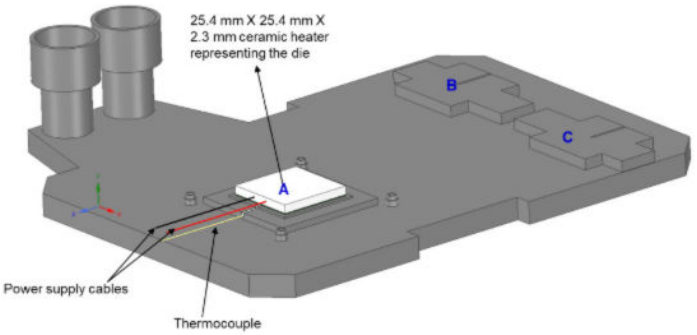


Figure 7: Instrumenting the cold plate; to provide better visualization, just one heater assembly is shown

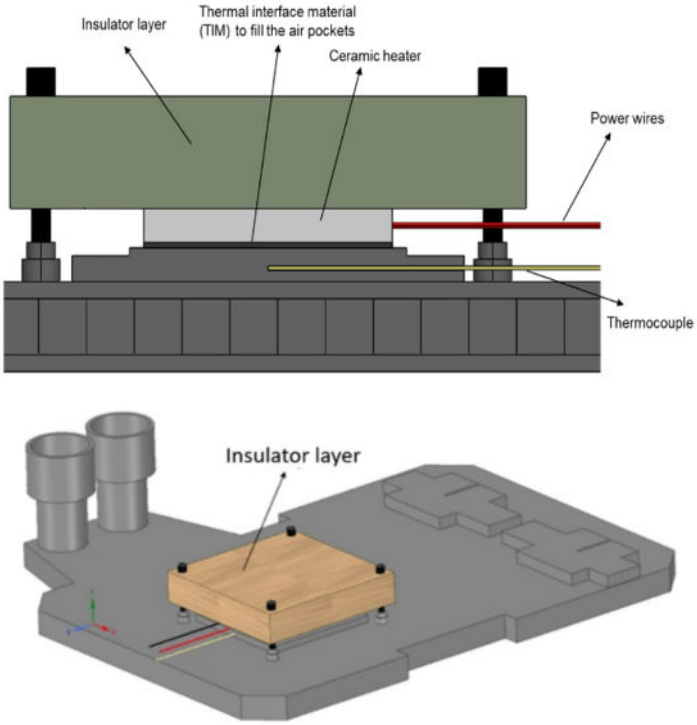


Figure 8: Heater module assembly

The schematic of the cooling loop is illustrated in *Figure 9*. All the sensors were calibrated prior to use.

Two thermocouples were placed inside the liquid loop to measure the coolant temperature upstream and downstream of the cold plate. To ensure accurate coolant outflow temperature readings, a custom-designed, 3D-printed static mixer was placed upstream of the thermocouples. Energy balance was achieved by calculating the heat generated by the heaters and comparing it to the energy extracted by the coolant in steady-state mode, resulting in a difference of less than 1% of the total heater energy. The orientation of the cold plate during testing did not impact hydraulic or thermal performance. *Figure 10* provides a snapshot of the experimental setup.

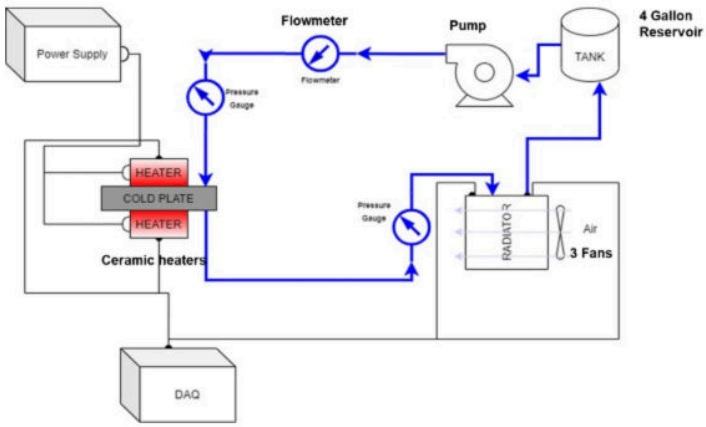


Figure 9: Simplified cooling loop schematic

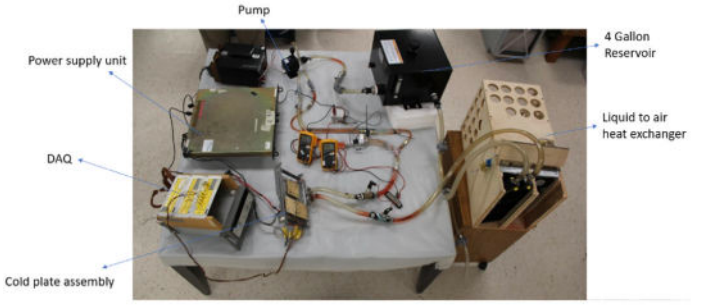


Figure 10: Overall experimental set-up

Validation of the CFD model

To validate the numerical simulations, test results for the hydraulic pressure drop across the cold plate are compared to simulations in *Figure 11*. The CFD model for hydraulic pressure loss correlated with the experimental data over a wide range of operating conditions. The minimum and maximum pressure loss of the unit were measured to be 0.4 kPa and 10 kPa at 0.4 and 6.4 LPM, respectively.

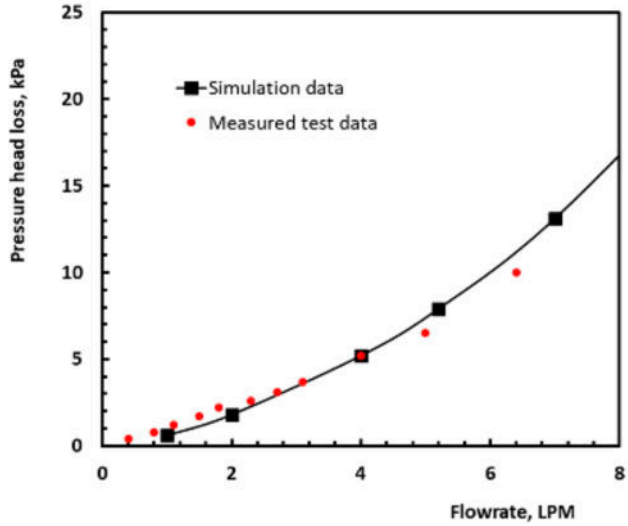


Figure 11: The pressure drop across the cold plate

	Pedestal temperature, °C					
	Case1 Coolant flow rate: 5.2 LPM			Case2 Coolant flow rate: 4 LPM		
	Simulation	Test data	T_{CFD}-$T_{experiment}$	Simulation	Test data	T_{CFD}-$T_{experiment}$
Chip A	45.3	46	-0.7	46	48	-2
Chip B	47.9	48	-0.1	48.9	50	-1.1
Chip C	47.3	48	-0.7	48.6	50	-1.4
Chip D	49.7	51	-1.3	51.1	54	-2.9

Table 1: Simulations vs. experiment results for coolant with an inlet temperature of 30°C, heat dissipation rate of 125.5 W per chip

The pedestal temperature test results of the four chips are compared to the simulation results in *Table 1*. This shows that the modeling results for the temperature map were consistent with the experimental data with a maximum temperature difference of less than 3°C.

The higher the flow rate, the lower the temperature of the chips. However, a more powerful pump is required to provide the coolant at higher flow rates. There is a tradeoff between the performance of the cold plate, the pump size, and the energy required to push the coolant through the cold plate. The liquid flow rate should be optimized to maintain a relatively low temperature rise from inlet to exhaust while also minimizing the impact on system pressure drop and pumping requirements.

Artificial Neural Network Model

The Artificial Neural network was organized in layers. The layers were made up of many interconnected *nodes* that contained *activation functions*. The number of the nodes in an input layer was set to 6 to include coolant inflow temperature, flow rate and thermal loads of four IC components. The input layer contained the values of the explanatory attributes for each observation. The hidden layers apply given transformations to the input values inside the network. For the cold plate assembly, two hidden layers were defined. The first and second layers included 12 and 6 hidden nodes, respectively. The response variables in the output layer included hydraulic pressure loss and the junction temperature for the four IC components. In this analysis, the temperatures of the pedestals beneath each heater module were used to represent the junction temperatures. This approach allowed for consistent temperature measurement relevant to the thermal management of the ICs.

The learning set was obtained by performing a set of parametric CFD simulations in which the coolant viscosity varied with temperature. Design of Experiments (DOE) was conducted, applying Latin Hypercube Sampling method in the design space. The coolant flow rate and temperature varied between 2 LPM to 7 LPM and 6°C to 44°C, respectively. The thermal load of each IC was set between 0 to 200 W. In total, the DOE included 140 simulation points, with 120 of them randomly selected as the learning set, to avoid overfitting, and the remaining 20 points were used to validate the generated ANN model. The predicted results for the hydraulic pressure loss across the cold plate and the four junction temperatures were compared to those from sample set. The correlation coefficient (R^2) for both the pressure loss and the junction temperatures was calculated to be over 0.98. The ANN results were found to correlate well with the validated CFD results. The ANN model can be used to calculate the hydraulic and thermal performance of the cold plate at any operating conditions within seconds without compromising accuracy.

To show case the capabilities of the generated ANN model, it was applied to estimate the thermal capacity of the cold plate, assuming an upper junction temperature limit of 90°C and a coolant inlet temperature of 40°C. As shown in *Figure 12*, the current cold plate can remove a maximum of 155 W and 179 W from Chip A at a coolant flow rate of 2 LPM and 7 LPM, respectively.

The power dissipation rate of Chip A increases the junction temperature of chip B, C, and D by a maximum of 7°C, 4.6°C and 6.6°C at 2 LPM and 1.6°C, 0.4°C and 1.6°C at 7 LPM, respectively. Assuming the maximum heat dissipation rate for Chip A, the cold plate can remove a maximum of 125 W and 156 W from Chip C at 2 LPM and 7 LPM, respectively.

The cold plate has the capacity to remove a maximum of 110 W and 156 W from Chip B at 2 LPM and 7 LPM, respectively, as-

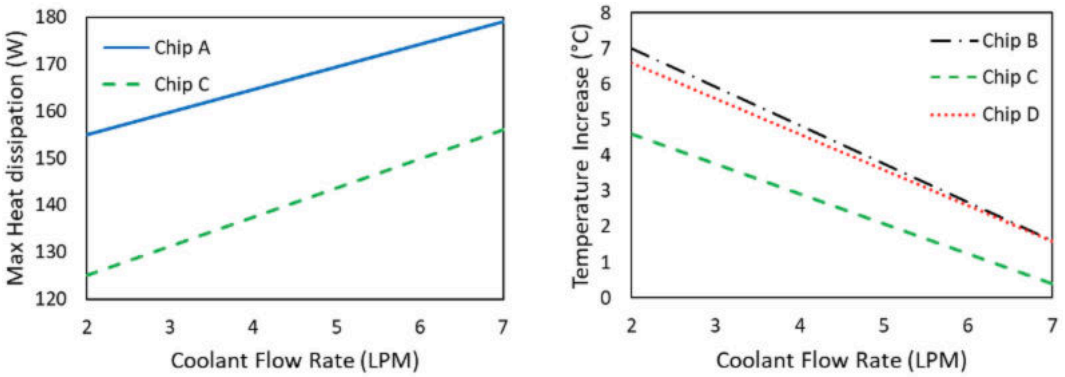


Figure 12: Maximum heat dissipation for Chips A and C at various coolant flow rates, along with the junction temperature increases for Chips B, C, and D due to Chip A's dissipation, based on an upper junction temperature limit of 90°C and a coolant inlet temperature of 40°C.

suming the maximum heat load from chip A and C. The maximum heat loads of Chip D were found to be similar to Chip B, 113 W and 155 W at 2 LPM and 7 LPM, respectively. The total thermal capacities of the cold plate were 503 W and 646 W at 2 LPM and 7 LPM.

Conclusion

It was shown that the developed artificial neural network model can represent the thermal and hydraulic characteristics of the cold plate. The results for pressure head loss and junction temperatures align closely with validated CFD results across a range of conditions. The model facilitates temperature estimation for critical components and evaluates cold plate performance under various coolant inflow conditions and power load distributions. The generated ANN model was used to identify performance limits of the cold plate. The developed model can accelerate cold plate design and development by expediting an evaluation of design feasibility and to conduct in-depth root cause analyses for various inputs and operating conditions.

References

[1] Azita Soleymani et al., "Hydraulic and Thermal Characteristic of a Double-Sided Cold Plate. Part 1: CFD Analysis", *Electronics Cooling Magazine*, Spring 2023.


Call for Authors
and Contributors!
Want to be a part of the next issue of Electronics Cooling? Have an article or blog post you'd like to write for Electronics-Cooling.com?
Let us know at
editor@electronics-cooling.com
electronics COOLING
www.Electronics-Cooling.com

Index of ADVERTISERS

Electronics Cooling
t: (484) 688-0300
e: info@electronics-cooling.com
w: electronics-cooling.com
page: 30

Henkel
w: BERGQUIST® GAP FILLER TGF 4400LVO
page: 6

CONTACT US

Lectrix
t: (484) 688-0300
e: info@lectrixgroup.com
w: lectrixgroup.com
page: 32


Break the same old pattern.
Problem First. Product Last.
Content | Data | Marketing Technology
LECTRIX®
Digital Marketing for the B2B Electronics Industry
1.484.688.0300 | info@lectrixgroup.com
www.lectrixgroup.com