Passive & Quasi-Active Approaches to Thermoregulation in the Characterization of Chiplet Based Devices in the Semiconductor Ecosystem

MP Divakar*, Ashok Kabadi*, Bryan Luippold**

***Ampere Computing LLC, ** ISC International**

Mesa, Arizona ● March 3–6, 2024

Contents

Drivers of Disaggregation in the Compute Ecosystem-1

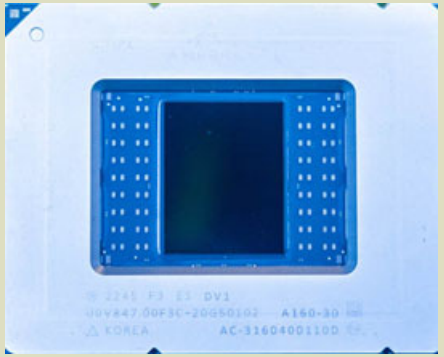
- Ever increasing compute demand ⟶ increasing core count in legacy x86 and ARM-based CPU architectures
- Computing nuances ⟶ GPU (graphics processing unit), NPU (neural processing unit) and TPU (tensor processing unit)
- Difficult to incorporate more compute cores, communication functions and tailored processors on the same substrate → chip reticle size limit
- Co-packaging processor types like CPU and GPU is still not feasible due to different junction temperature limits

Drivers of Disaggregation in the Compute Ecosystem-2

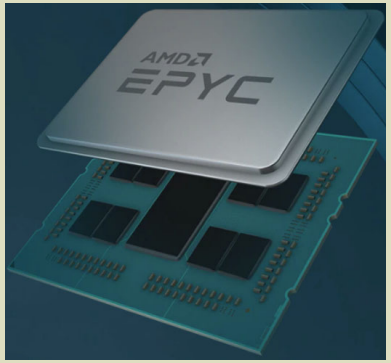
- Evolving artificial intelligence applications like ChatGPT, machine learning, inferencing, etc. ⟶ CPU developers seek external partners for specialized functional chiplets
- Emerging Universal Chiplet Interconnect Express (UCIe) standard for die-to-die communication ⟶ new solution providers, shorter development time
- Disaggregation ⟶ new era of **packaging** with **high I/O count**, **large body sizes** and **socketable** land grid array (**LGA**) as preferred attributes

Examples of Disaggregation in the Compute Ecosystem

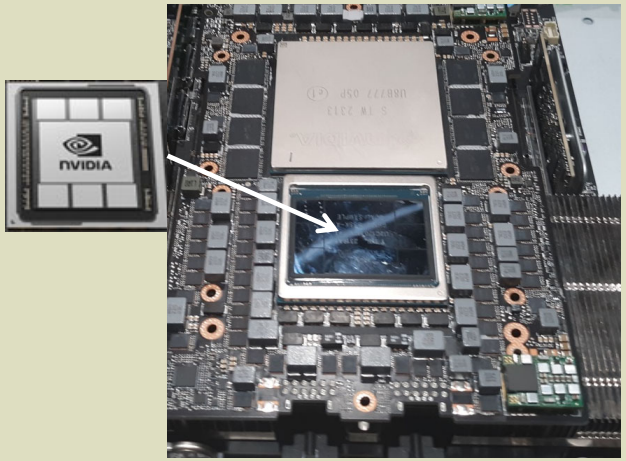
- AMD EPYC®, Ampere Computing's AmpereOne®, Nvidia's Grace® CPU in the Grace-Hopper® superchip configuration, Intel's scalable Xeon®, etc.
- Disaggregated CPUs offered in lidded (heat spreader) as well as lidless or partially lidded packages



Partially Lidded AmpereOne®



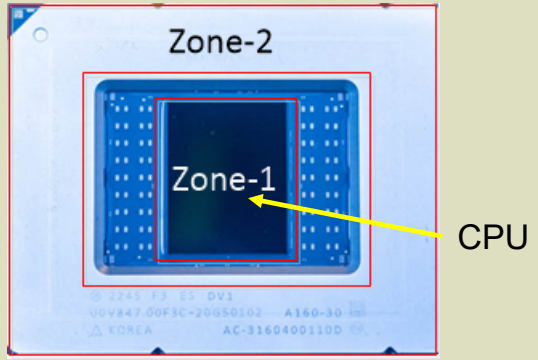
Fully Lidded AMD EPYC®

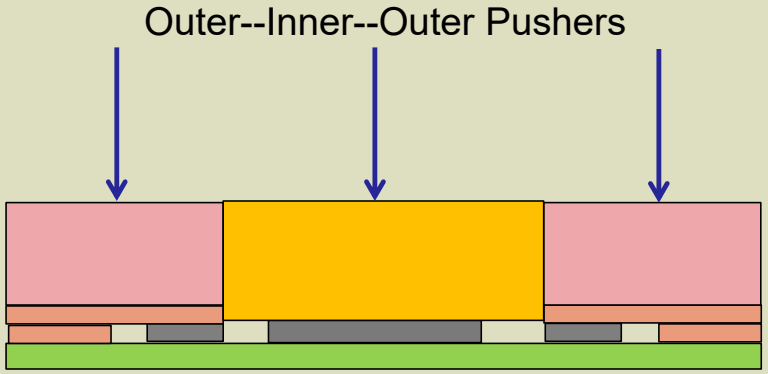


Lidless Nvidia Grace CPU

Challenges in Thermal Regulation of Lidless & Partially Lidded CPU Packages

- Thermal regulation **stack:** chiller and controller for TCUs, passive thermal lid or active temperature control unit (**TCU**), device under test (DUT), socket, validation/test PCB
- For partially lidded packages, mechanically- and thermally-interfacing to two datums is challenging
- Socketed CPU → adequate force to maintain lowest contact resistance –e.g. 35gm/contact, ~6000+ contacts
- Ergonomics → size and weight

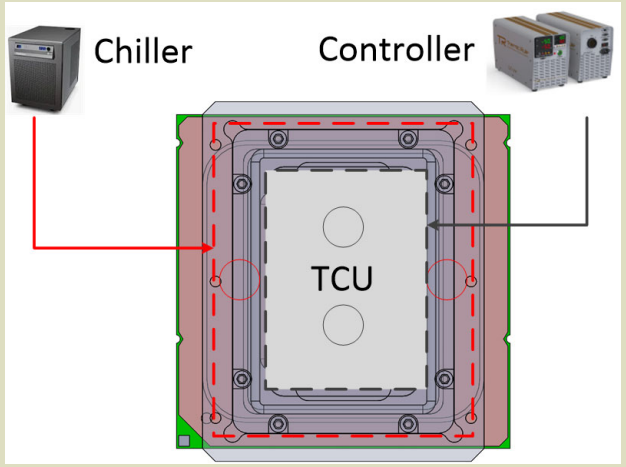




Partially Lidded CPU

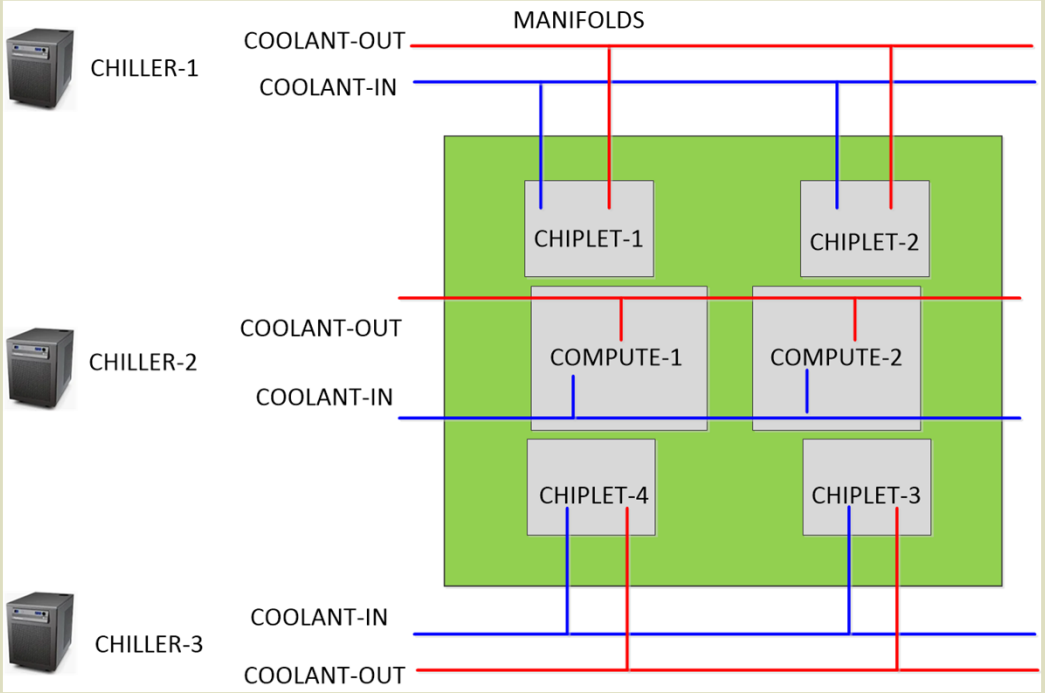
Approaches to Thermal Regulation in Validation Systems

- **Active**: TCUs provide temperature regulation above or below the chiller set temperature to meet test temperature
- TCU architecture: air pressure- or electrically actuated thermal head, a **controller** to operate thermoelectric coolers (TEC), a chiller and a software interface
- **Passive**: temperature control accomplished using a heat exchanger
- Targeted coolant flow ⟶ reduced thermal gradients across surfaces in contact with the CPU
- Passive thermal solutions ⟶ a fraction of the cost of TCU; lower maintenance



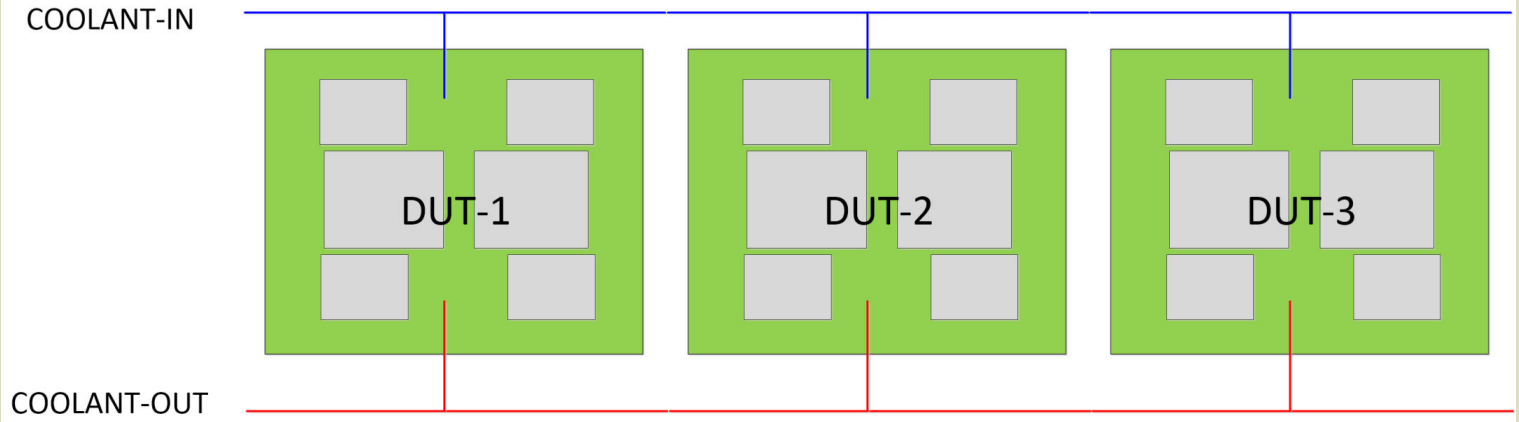
Design Approach for Multizone Passive Thermal Heads

- Thermal design power (TDP) and the number of independent control zones
- Generate design concept for the heat exchangers in each zone
- Computational fluid dynamics (CFD) simulations to estimate the temperature rise at thermal interfaces of each zone
- Bracketing studies to estimate the effect of chiller temperature and coolant flow rate on temperature rise (TRise)



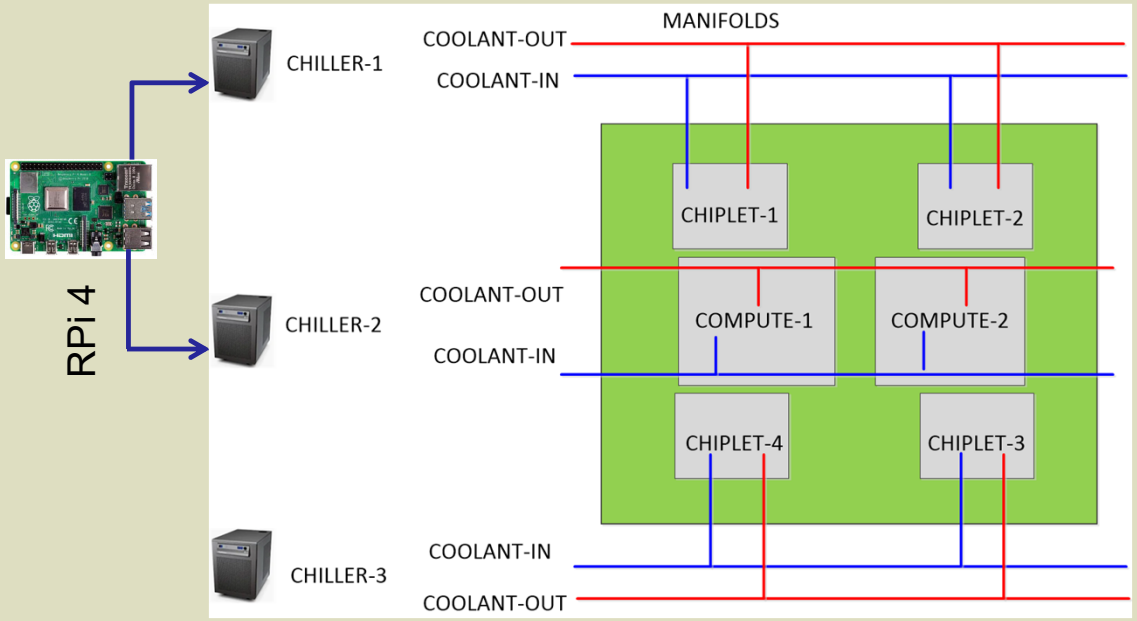
Scaling Multizone Thermal Control Approach for Multiple Validation Systems

- The number of validation systems that can be supported by each chiller for each zone depends on the TDP and test temperatures vs. heat removal capacity of the chiller
- Each chiller can support multiple validation systems by using input and return manifolds

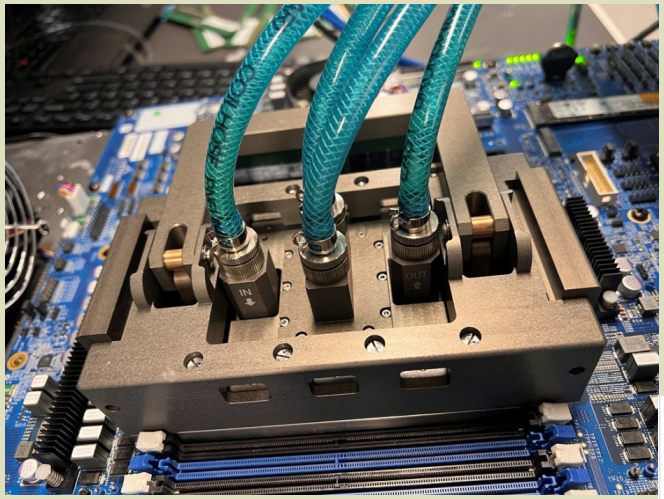


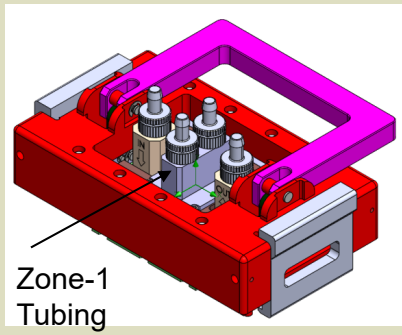
Building A Quasi-Active Thermal Regulation System

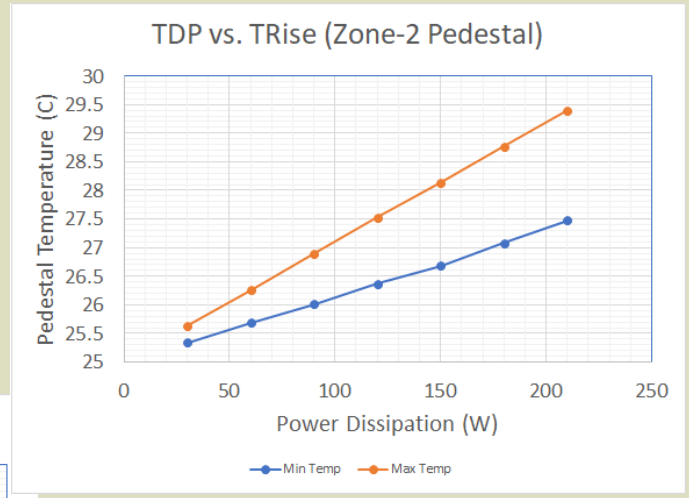
- Chiller temperatures are set based on the DUT temperatures for each zone
- The TDP vs. TRise for each zone is calculated by simulation (next slide)
- To set nearly isothermal boundary on the DUT, the chiller temperatures are offset as needed
- Raspberry Pi 4 is used to vary the chiller setting as needed to achieve a quasi-active thermal control system

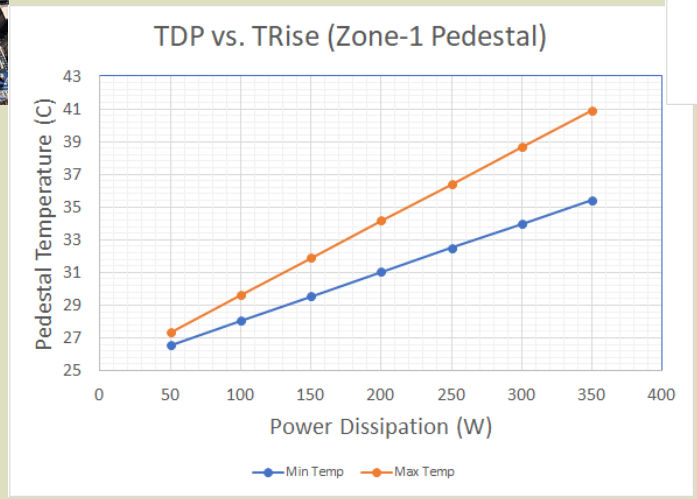


Example : Dual Zone Thermal Head in Validation PCB







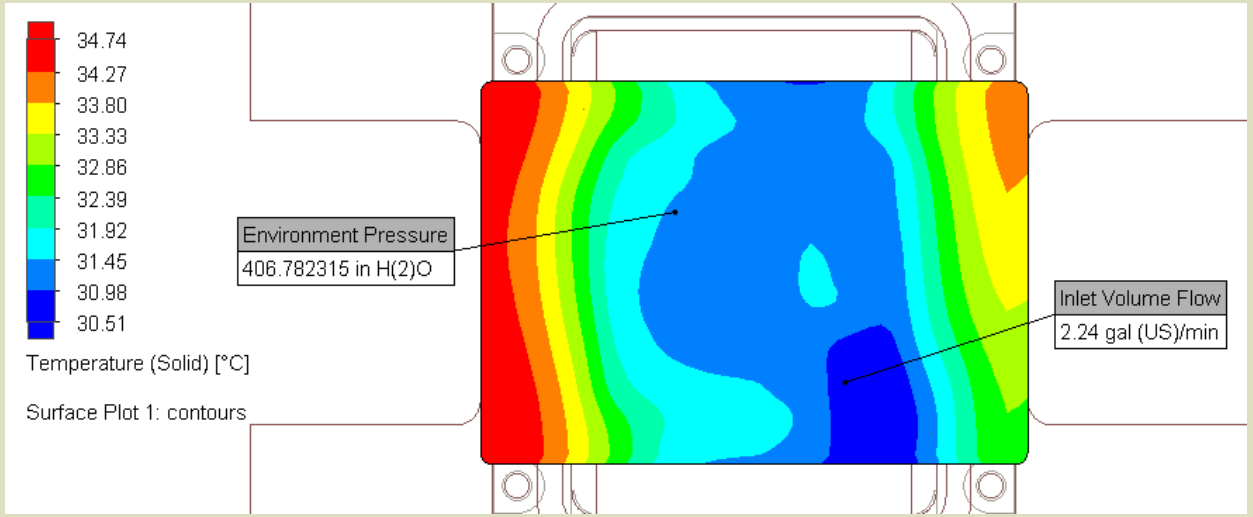


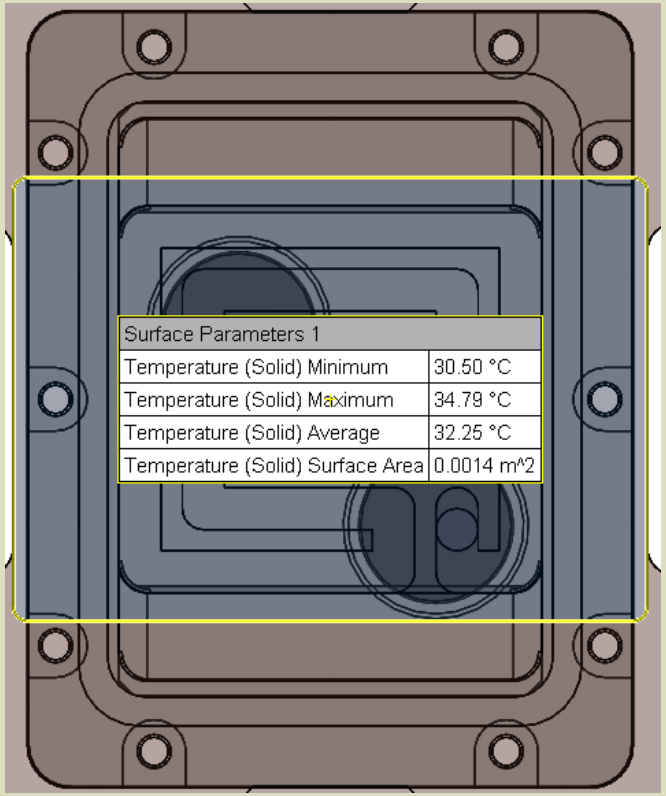
For temperatures <10C, a purge box is used to prevent condensation (not shown in the image above)

Pre-optimization sensitivity study of TDP vs. TRise in Zone-1 and Zone-2 by CFD simulation

Example: Dual Zone Thermal Head in Validation PCB

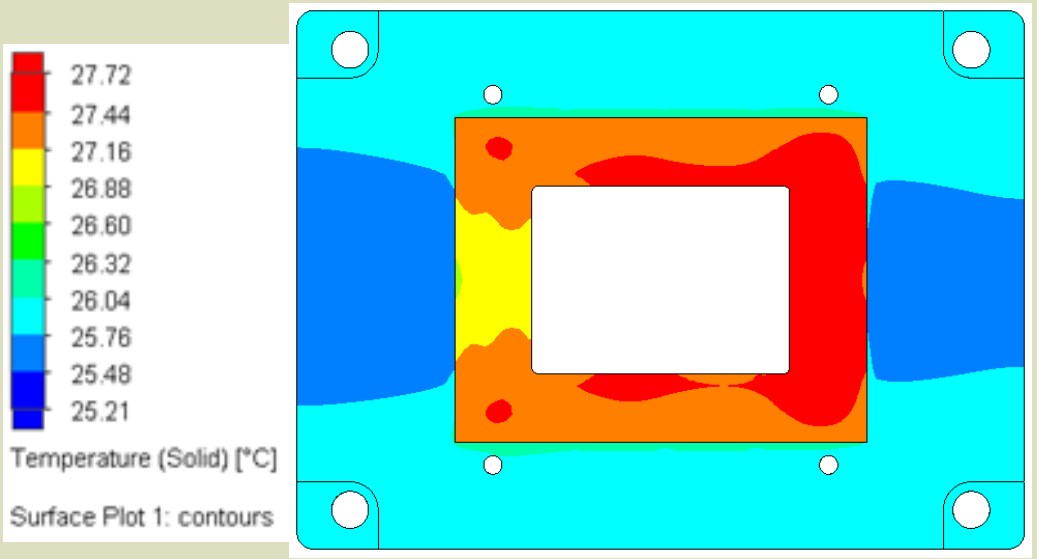
- **Zone-1** pedestal gradient less than 3.25C. This can be further reduced by design optimization of that zone.
- For TDP of >300W, the temperature rise above chiller setting was ~10C.





Example: Dual Zone Thermal Head in Validation PCB

- **Zone-2** pedestal surface gradients less than 1.25C for a TDP of >100W. This gradient can be further reduced by design optimization.
- For TDP of >100W, the temperature rise above chiller setting was ~2.5C
- Various TDP cases and test temperatures studied → chiller offset temperatures between zone-1 and zone-2 from 2C to 10C

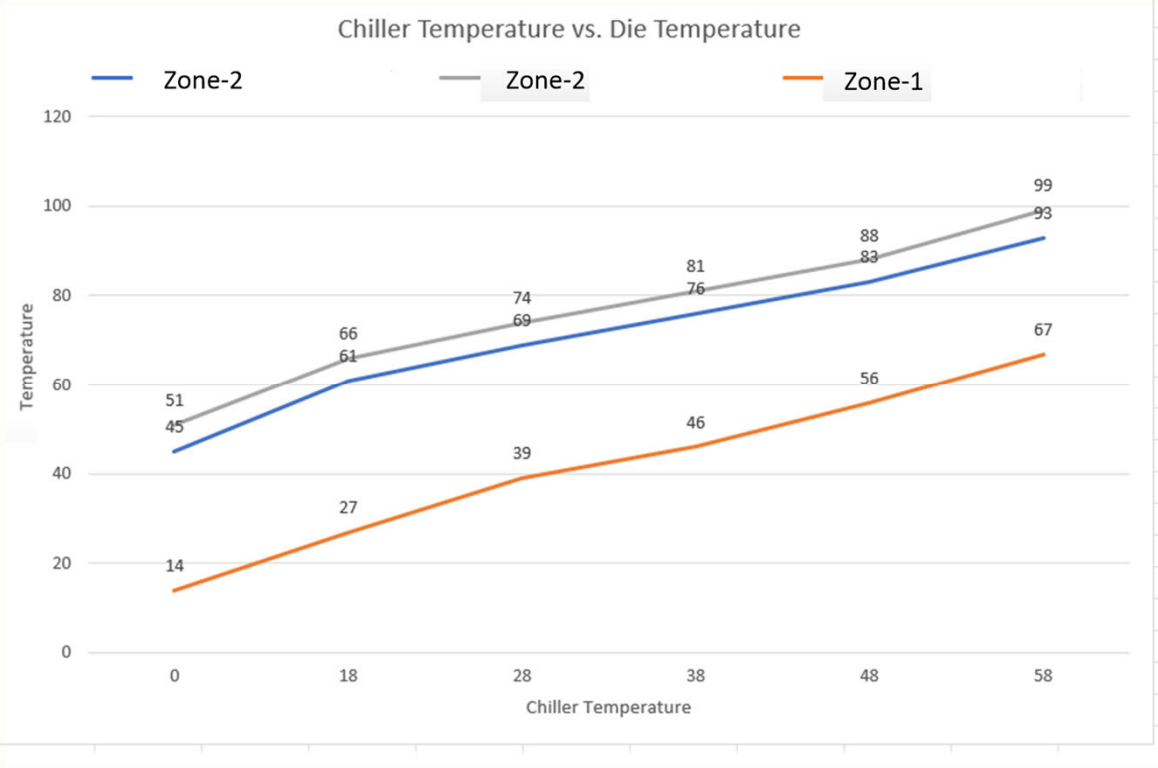


Chiplet Temperature Measurement: Zone-1 (Low) & Zone-2 (High)

- Graph shows measured temperatures in the two zones for various chiller settings

-The test goal was to make chiplets in zone-2 reach the maximum temperatures with minimal work loads

-Zone-1 was set high, the chip in a quiescent state where the temperature build up above the chiller temperature is ~10C



Conclusions

1. A design approach for low-cost multi-zone passive thermal hat for application in validation systems is presented. A dual zone passive thermal hat was designed as a test case using CFD simulations and optimized.
2. The passive thermal hat was tested at varying chiller temperatures, it maintains stable surface temperatures for the test cases investigated.
3. It was possible to reach nearly 100C in one of the zone-2 with workloads
4. In zone-1 with the compute chip in a quiescent state, the thermal hat maintains a constant thermal gradient under variable chiller temperature setting.
5. The passive thermal hat solution costs a fraction of the traditional TCUs and requires very little maintenance.
6. Some measure of automation can be accomplished by connecting a single board computer like Raspberry Pi to control the chiller temperatures remotely.

References

1. “Proprietary Vs. Commercial Chiplets” by Ed Sperling, Semiconductor Engineering, December 19-th, 2023: https://semiengineering.com/proprietary-vs-commercial-chiplets/

2. “Test Cell Thermal Solution" by Gianluca Lombardi, BiTS 2016

3. "Device Characterization Over Temperature at the Board Level" by Barry Johnson, BiTS 2016

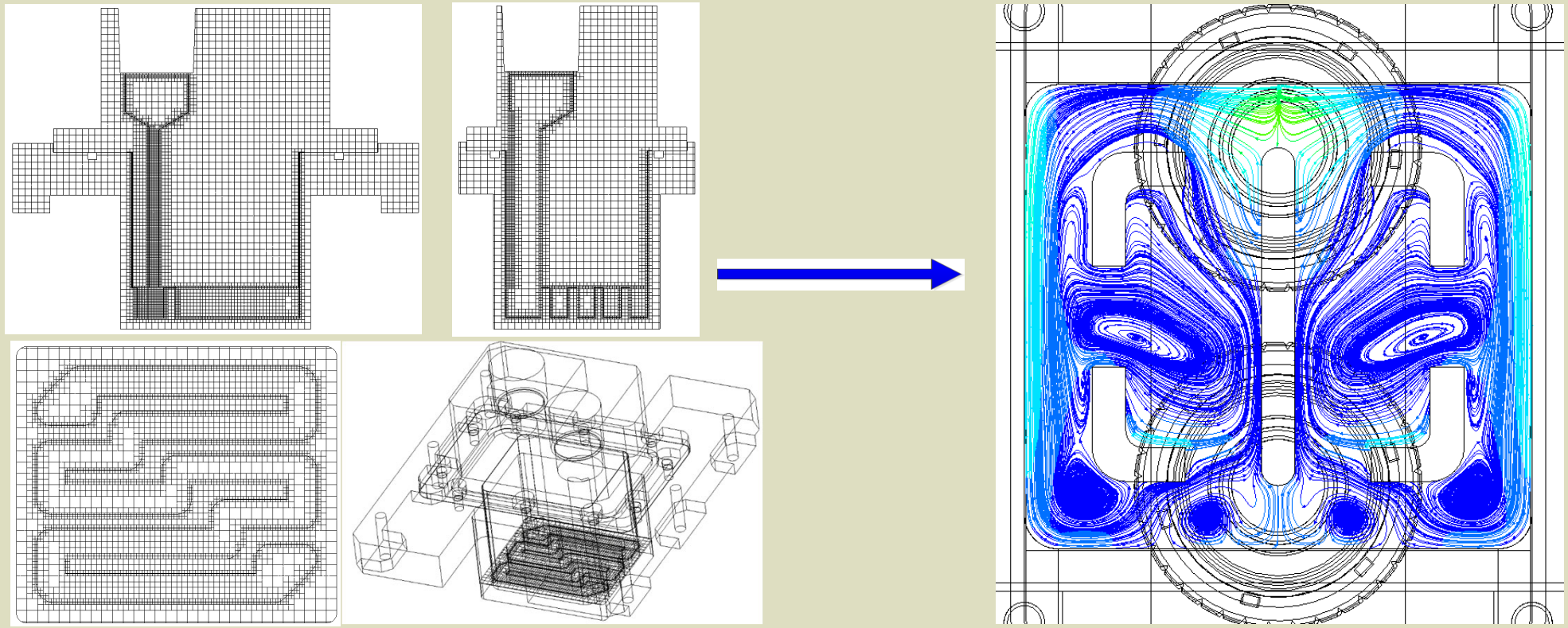
4. “IC Test Thermal Management” by Quynh Nguyen, TestConX 2020

5. “System Level Test for Automotive Devices -A Thermal Perspective” by See Jean “Sid” Chan & Wee Tick Lo, TestConX 2020

6. “Thermal Challenges & Cooling Methods for High Power Electronics” by Suhail Azim, Kenny Lee & PY Goay

7. “Closer Tolerance Thermal Management at the Device-Under-Test” by Bert Brost & Mehdi Attaran, 2019

CFD Can Be Fun!



COPYRIGHT NOTICE

The presentation(s) / poster(s) in this publication comprise the Proceedings of the TestConX 2024 workshop. The content reflects the opinion of the authors and their respective companies. They are reproduced here as they were presented at the TestConX 2024 workshop. This version of the presentation or poster may differ from the version that was distributed at or prior to the TestConX 2024 workshop.

The inclusion of the presentations/posters in this publication does not constitute an endorsement by TestConX or the workshop's sponsors. There is NO copyright protection claimed on the presentation/poster content by TestConX. However, each presentation / poster is the work of the authors and their respective companies: as such, it is strongly encouraged that any use reflect proper acknowledgement to the appropriate source. Any questions regarding the use of any materials presented should be directed to the author(s) or their companies.

www.testconx.org