



Confined Direct Two-phase Jet Impingement Cooling with Topology Optimized Surface Engineering and Phase Separation Using Additive Manufacturing

Dr. Tiwei Wei^{1,2}

¹University of California, Los Angeles; ²Purdue University



Project Vision – To address the instability, boiling crisis and reliability of confined two-phase jet impingement cooling, our solution includes new algorithms for topology optimization of the cooling structure, novel on-chip direct printing methods for laser powder bed fusion of multi-porosity wicks, and an additively manufactured multi-input/multi-output fluid distribution manifold.

Team



Prof. Tiwei Wei
UCLA; Purdue University (ME)



Prof. Justin Weibel
Purdue University (ME)



Prof. Bahgat G. Sammakia
Binghamton University (ME)



Prof. Scott Schiffres
CHIPADD; Binghamton
University (ME)



Prof. Srikanth Rangarajan
Binghamton University
(SSIE)



Dr. Ryan Enright
Seguente



Dr. Raffaele Luca Amalfi
Seguente



Dr. Arad Azizi
Solstice

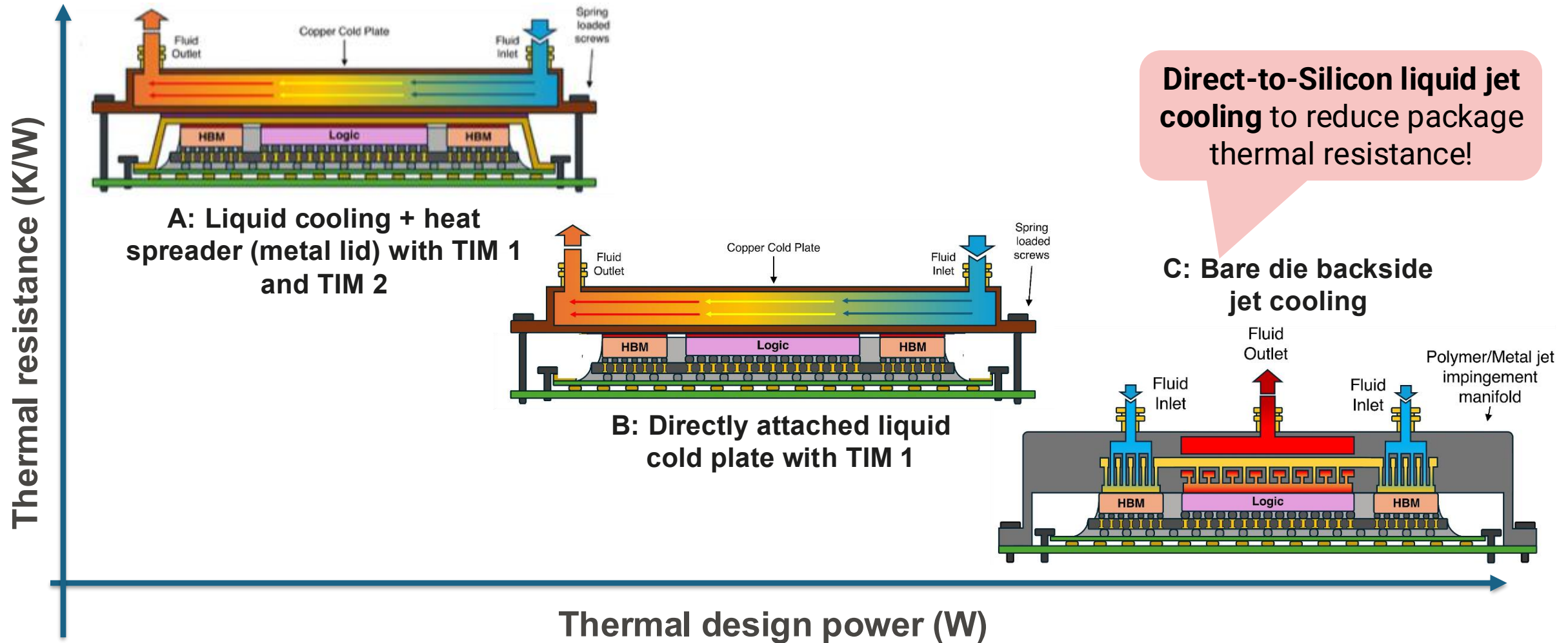


Dr. Ahmad Gharaibeh
Vertiv

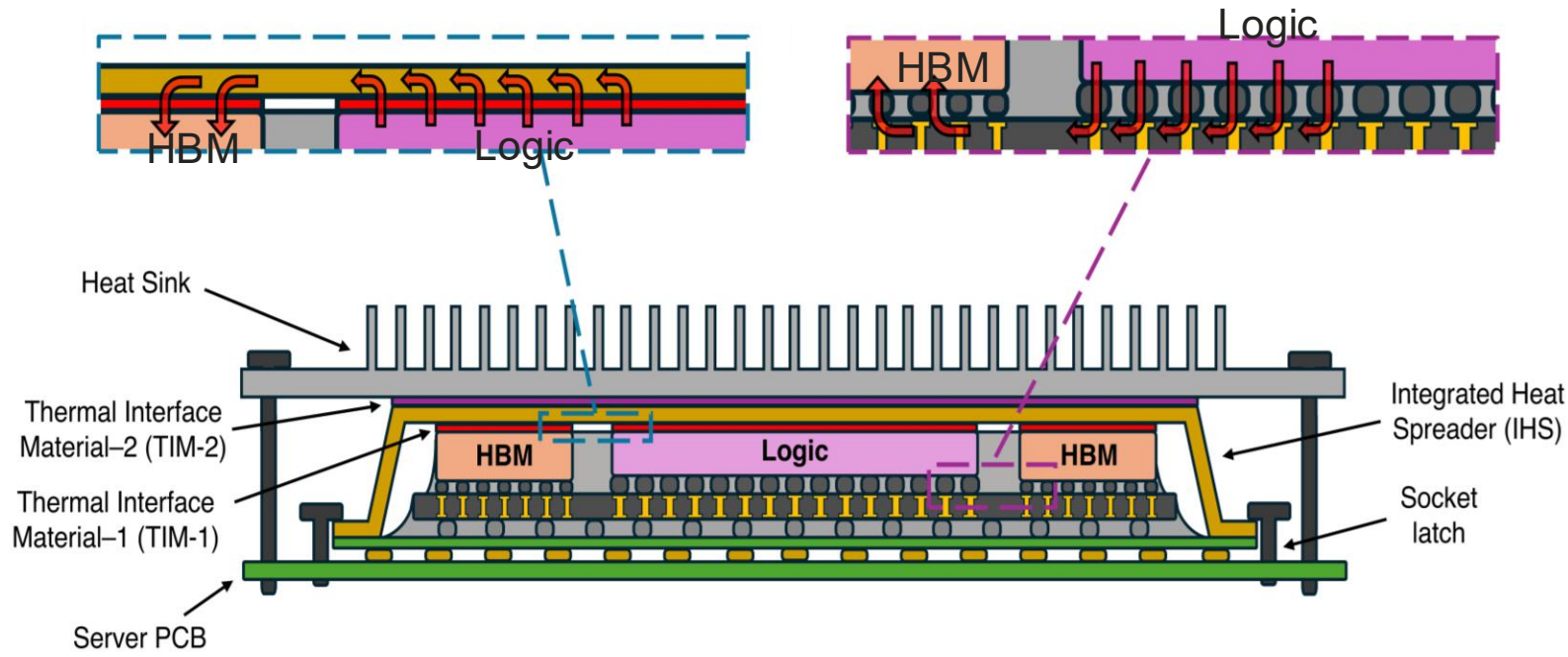


Dr. Yaman Manaserh
OMNICOOL

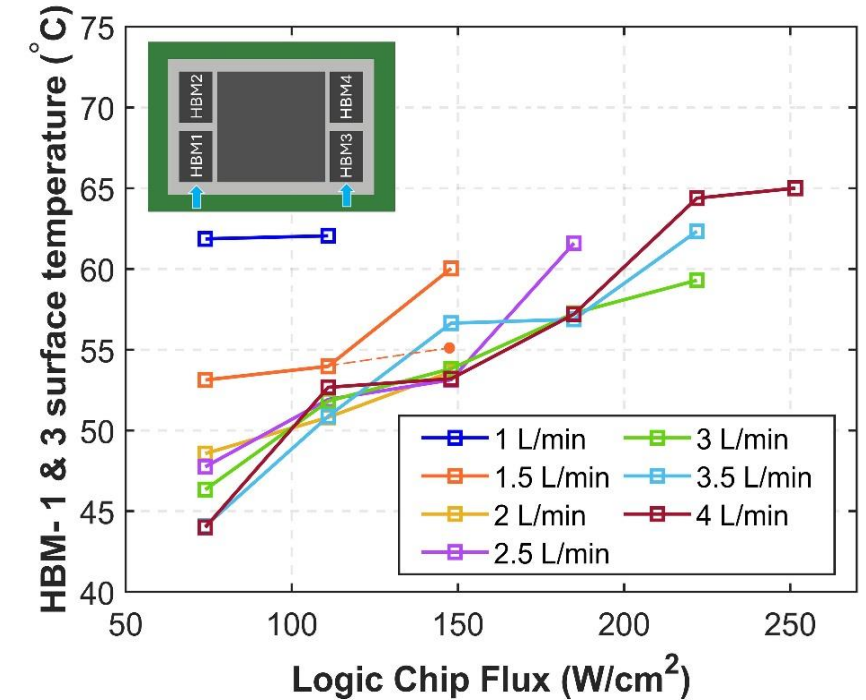
Motivation: Direct-to-Silicon Liquid Cooling to Eliminate the TIMs



Motivation: Direct-to-Silicon Liquid Cooling to Minimize Thermal Crosstalk



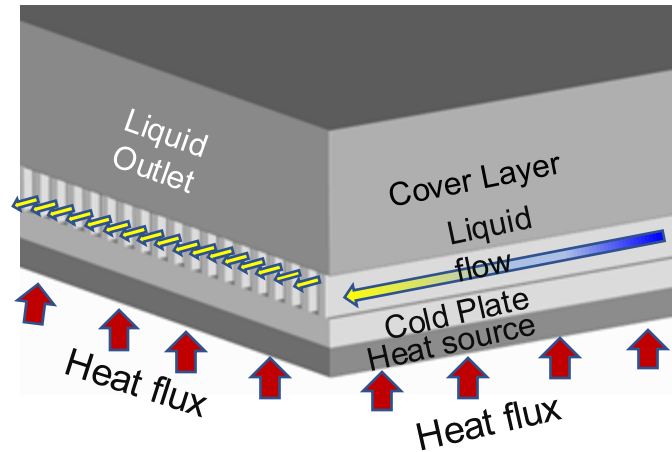
- As the GPU power keeps increasing towards 2.5 kW and 3 kW, the thermal coupling between GPU and HBM will become significant;
- Direct-to-silicon can effectively cool the GPU and HBM separately, especially for Megawatt Rack.



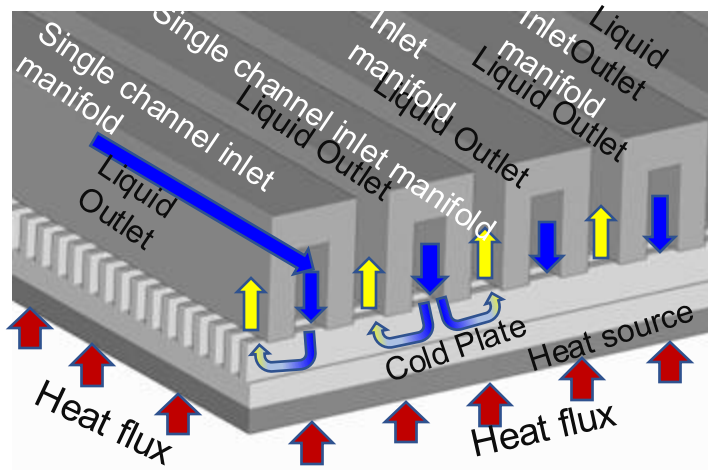
Patel, A., .. & Wei, T. (2025). Int. J. Heat Mass Transf, 244, 126978.

Problem/Set Up: Needs a More Cost-effective, Energy Efficient Solution

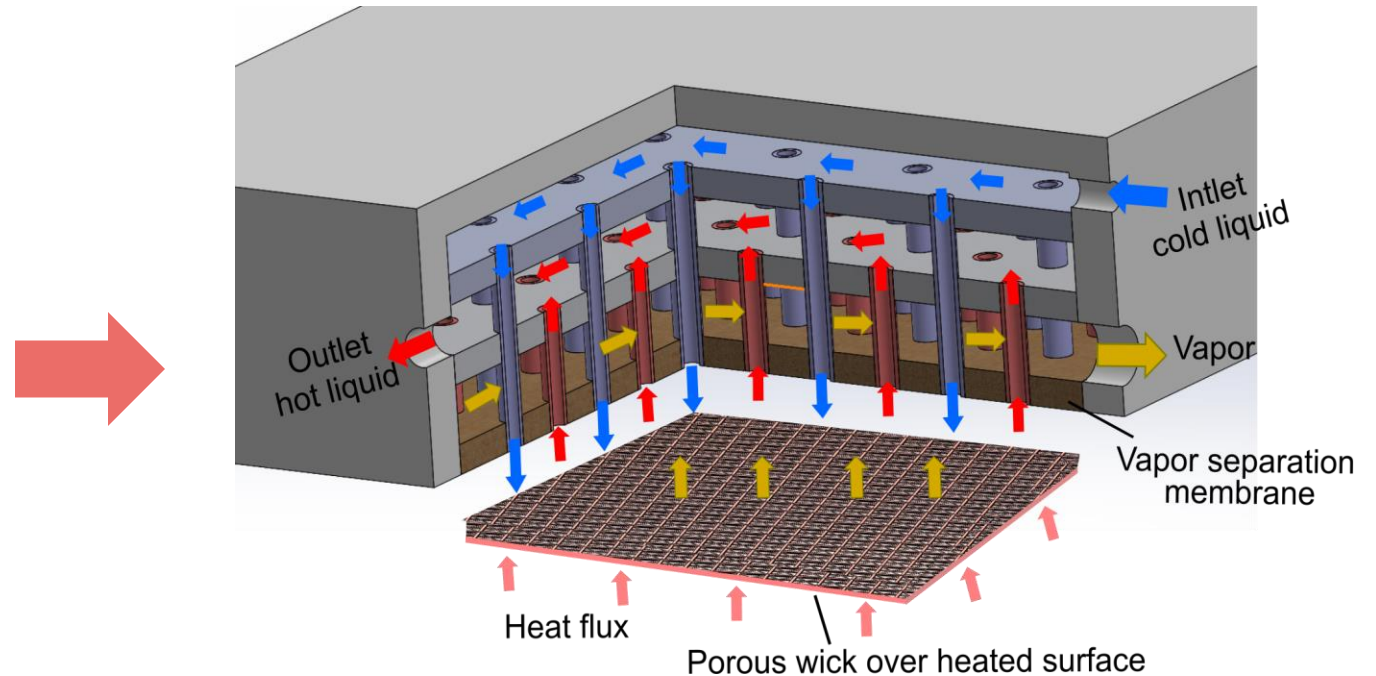
Conventional 2D microchannel cooler



3D manifold microchannel cooler

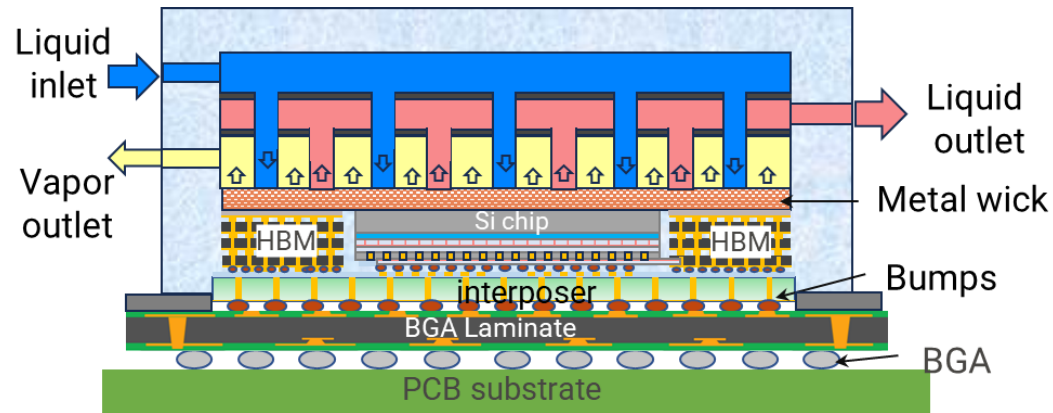


Direct-to-Si jet impingement cooling with distributed inlet-outlet nozzles

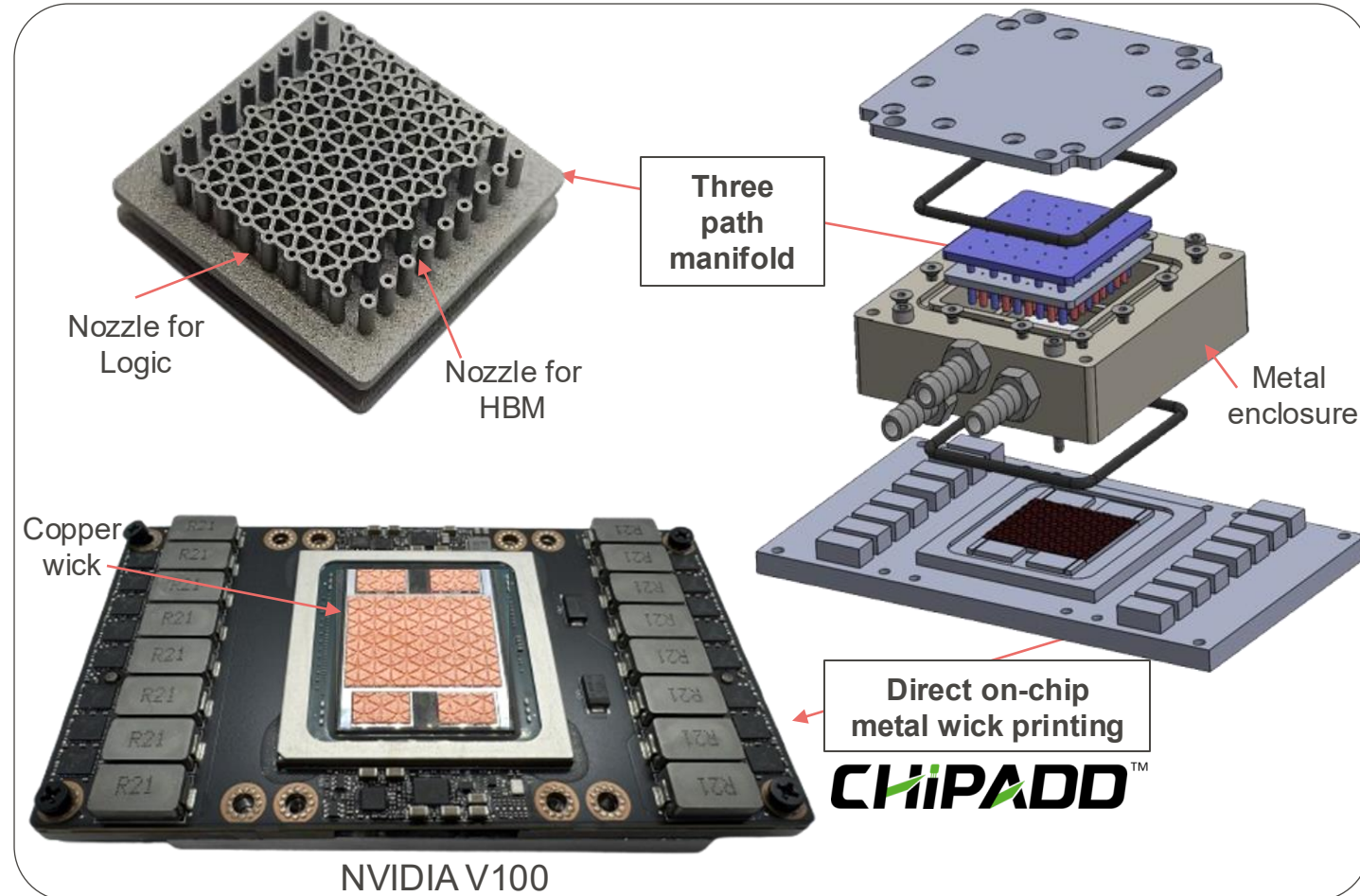


- Micro-channel approach introduces risk and cost by requiring lithography and etching of the back-side silicon
- Jet impingement is more scalable to large chip sizes with lower pressure drop and more uniform temperature

Our Solution: Direct-to-Silicon Jet Cooling: Two-Phase R1233zd(E)



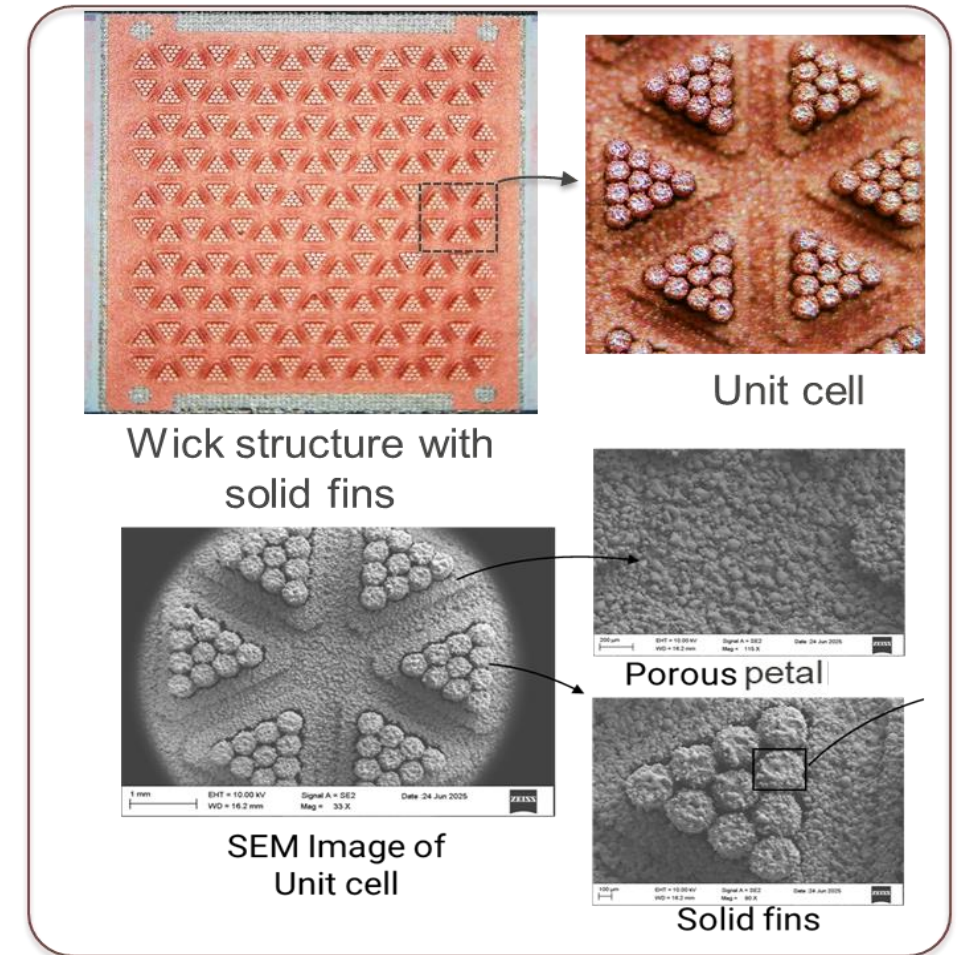
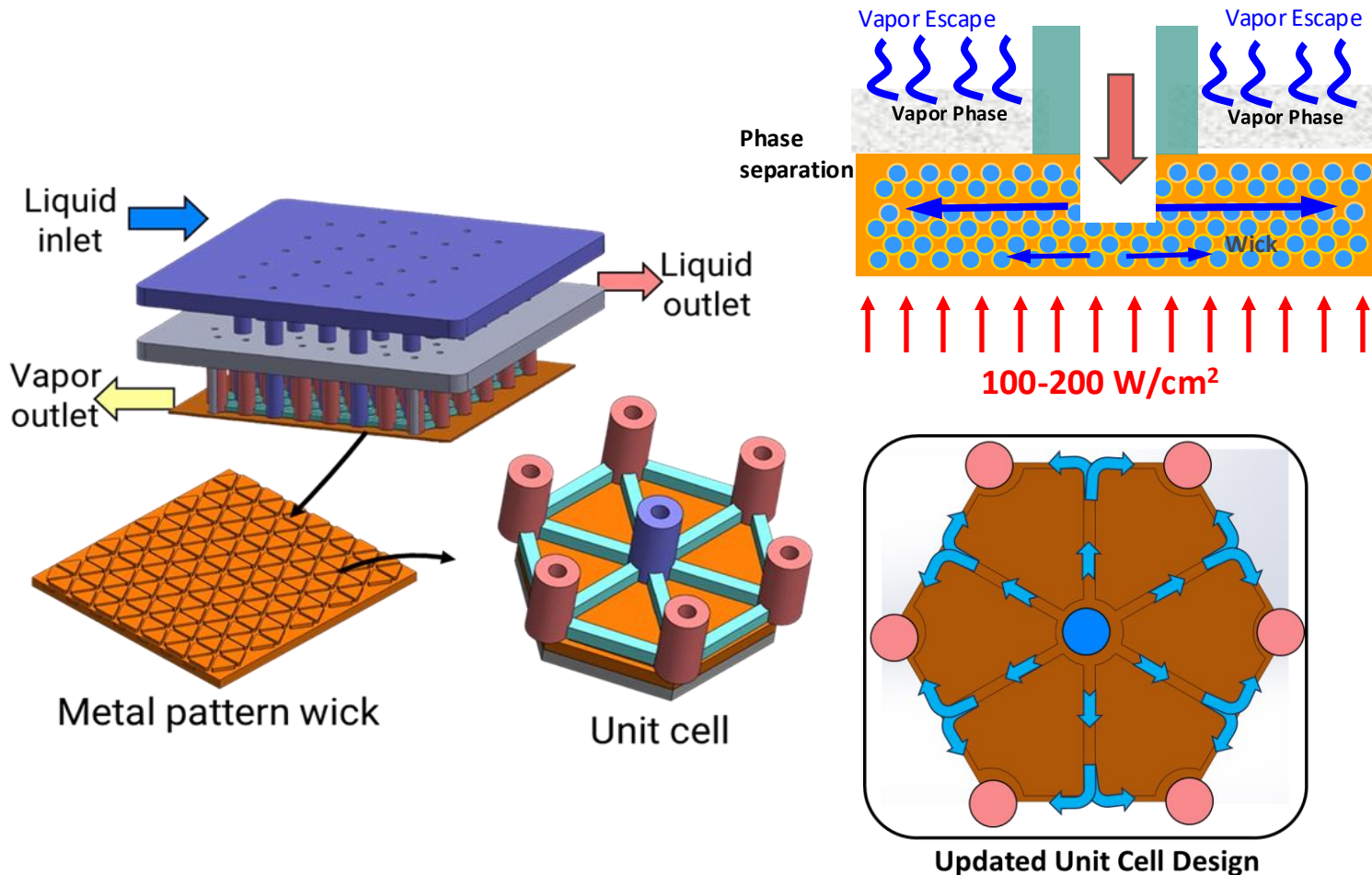
- Novel liquid and vapor separation improvement method for two-phase jet impingement
- Novel on-chip direct printing method
- Ultra compact and thin lid integrated impingement multi-jets manifold for on-chip cooling



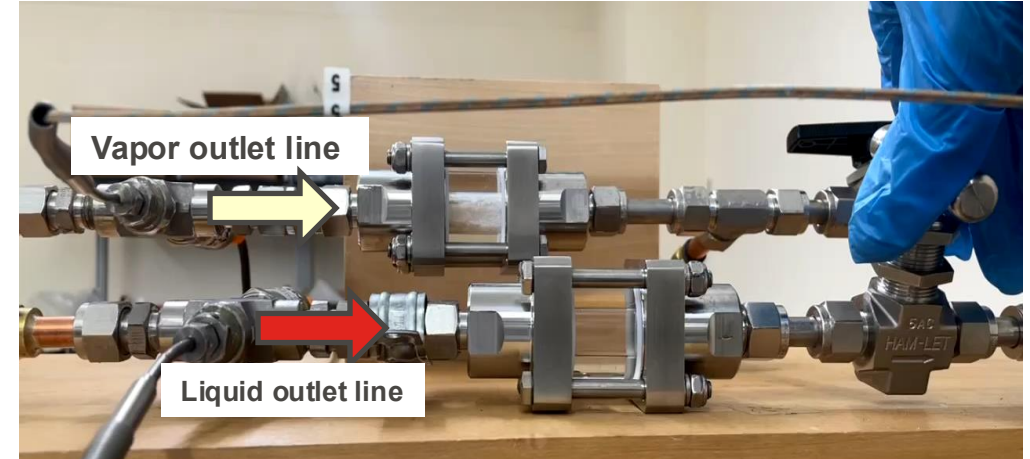
- Purdue patented: Novel Liquid and Vapor Separation Improvement Method for Liquid Jet Impingement
- Purdue patented: Ultra Compact and Thin Lid Integrated Impingement Multi-jets Manifold for on-Chip Cooling
- Binghamton patented: Direct-on-Chip Metal Printing

Technology Description: Surface Engineering and Phase Separation

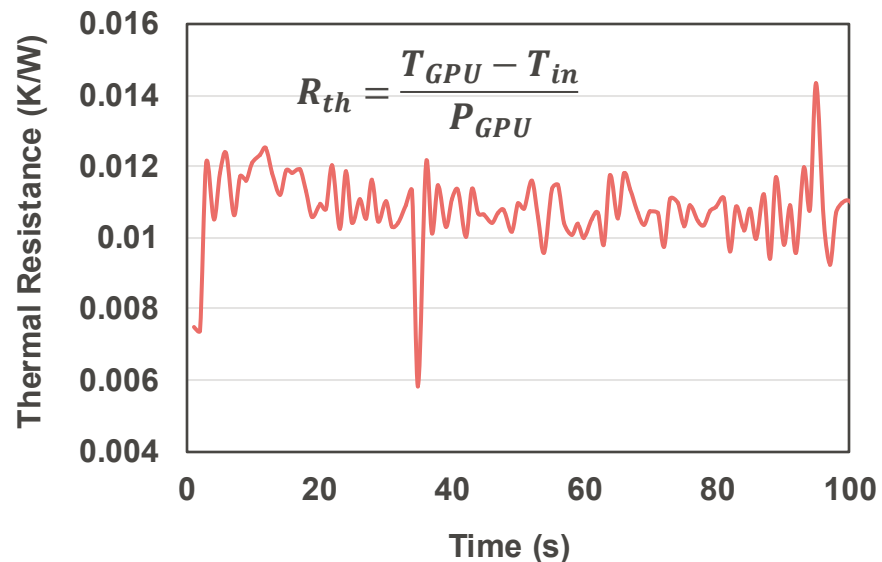
Confined direct two-phase jet impingement cooling with R123zd(E) as working fluid



System Integration Results and Midterm Milestone

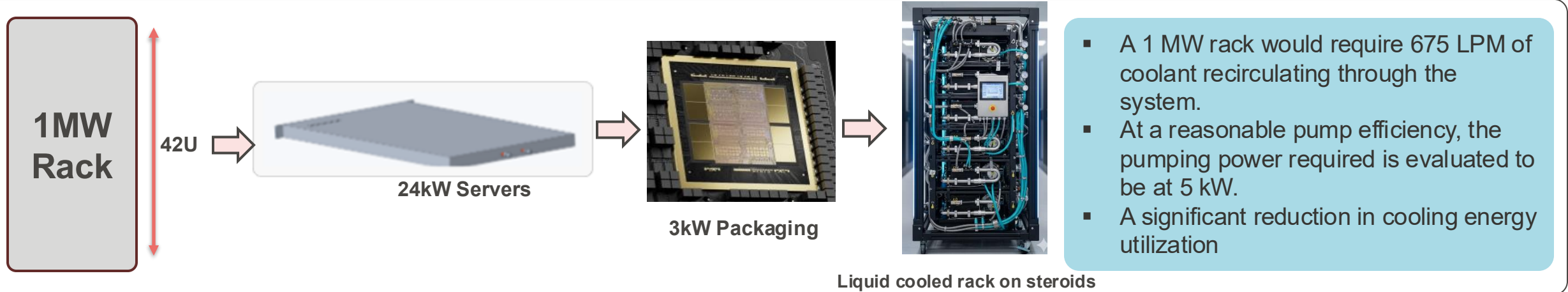


Integration of manifold with wick printed V100 GPU



- Successfully tested wick printed V100 GPU with direct on chip two-phase liquid jet impingement, which delivers 300 W of TDP.
- Successfully demonstrated functionality of the V100 GPU after printing of the wick structure
- Performed reliability testing for more than 1000 hrs of continuous testing.
- System availability: 98.88%

The Megawatt Rack

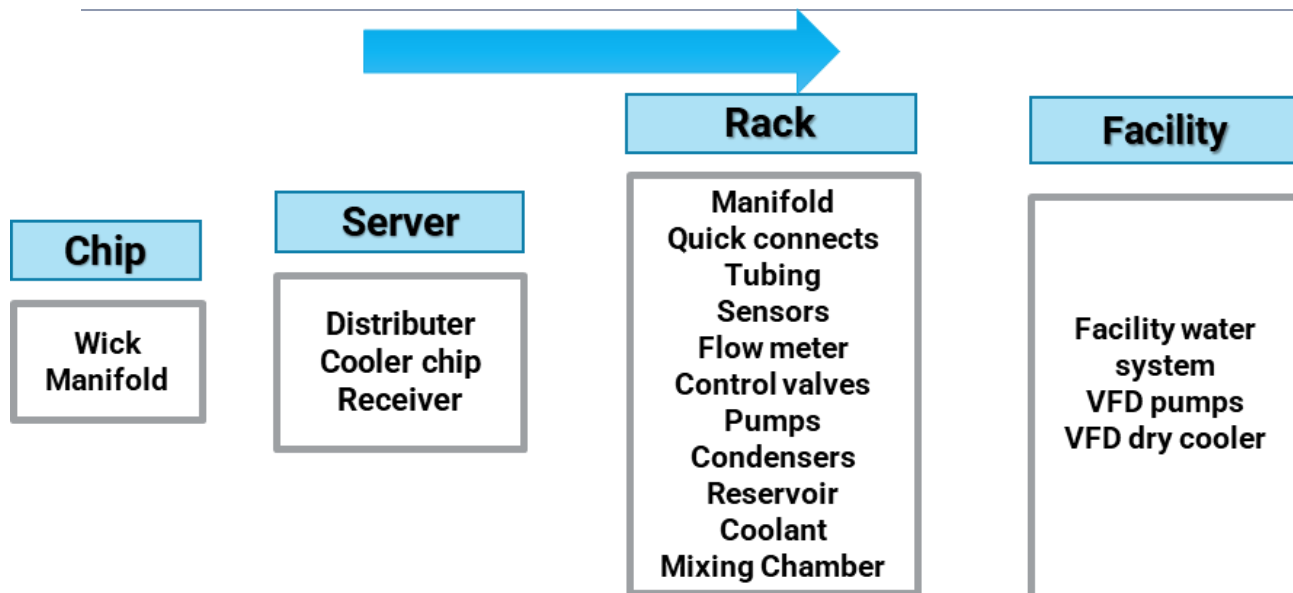


Integration of cooling solution on 1U server

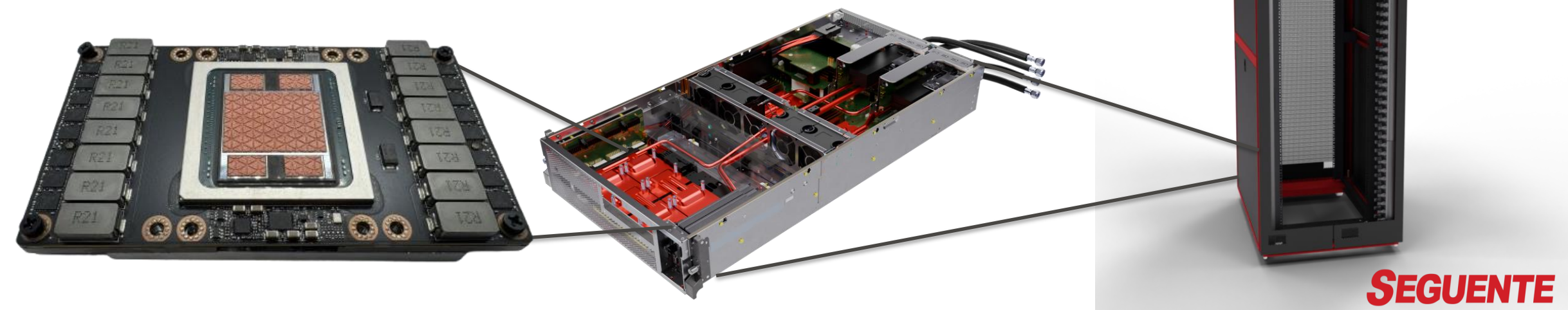
Scale to the Megawatt rack: challenges

- Extremely high vapor volume outflow for two-phase operation in MW racks.
- The recommended path forward is to utilize a vapor separator cooling approach for MW server racks, which we will demonstrate to enhance enhanced system stability in our rack-level testing.
- For the CDU, its control capabilities—particularly flow rate and temperature regulation must be evaluated across a wide range of flow conditions to address the rising CPU power demands of the 1 MW rack.

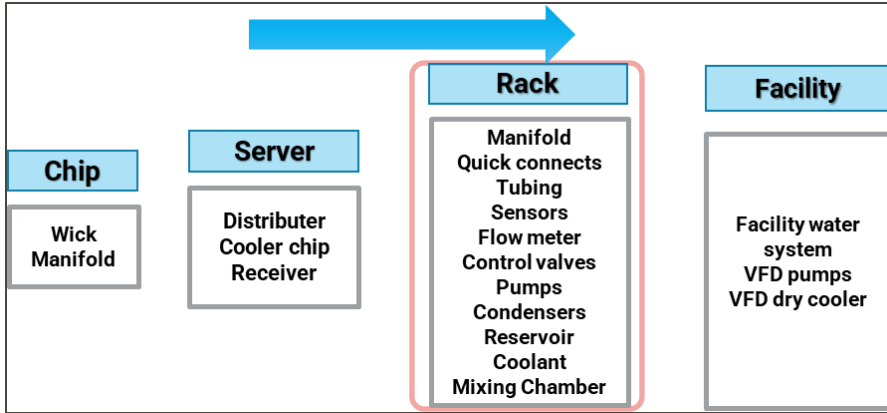
What is Next: Passive and Pumped Rack Cooling System Integration



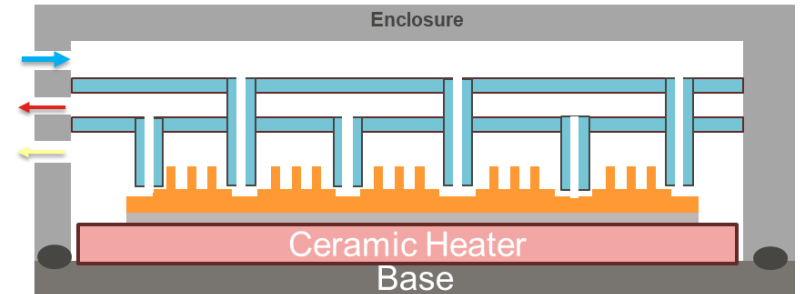
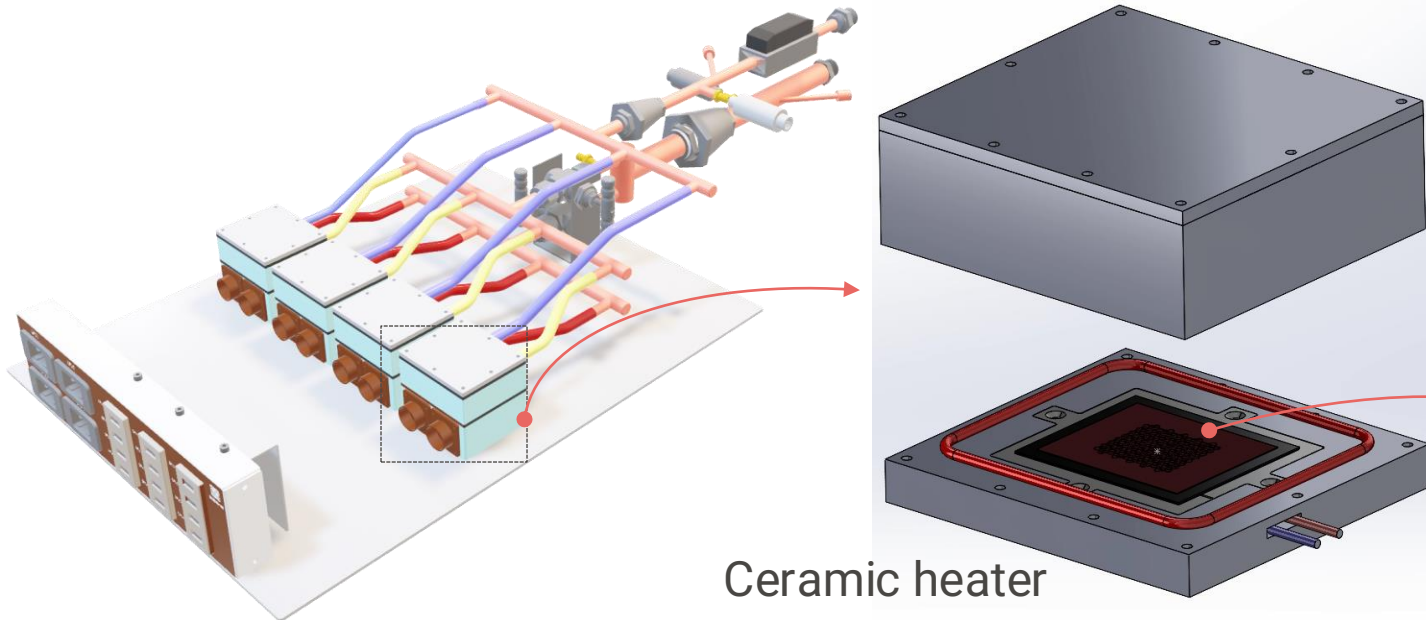
- Transfer direct printing process to Nvidia H100
- Deploy direct-on-chip cooled GPUs in server
- Demonstrate in passive and pumped rack cooling system



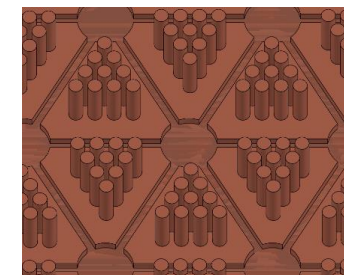
What is Next: Rack Level Demonstration



- The heater is designed to replicate the NVIDIA B200 GPU.
- The heater can go up to 1.5 kW of heat load.
- An NVIDIA thermal test vehicle is to be used that contains four heaters, making a total heat load of 6 kW per TTV.
- The rack demonstration is assembled to test a heat load of 12 kW in a rack CDU.

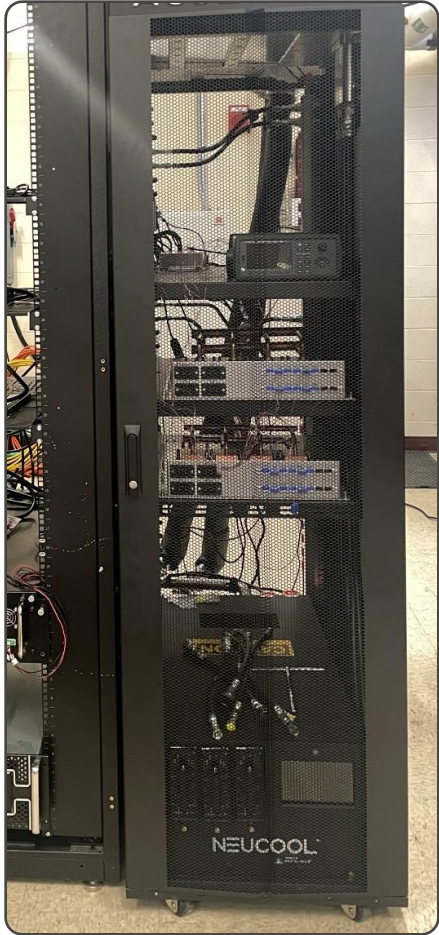


Heater assembly schematic



wick structure

What is Next: Rack Level Demonstration



Two-phase in rack CDU system
(CDU contributed by VERTIV)



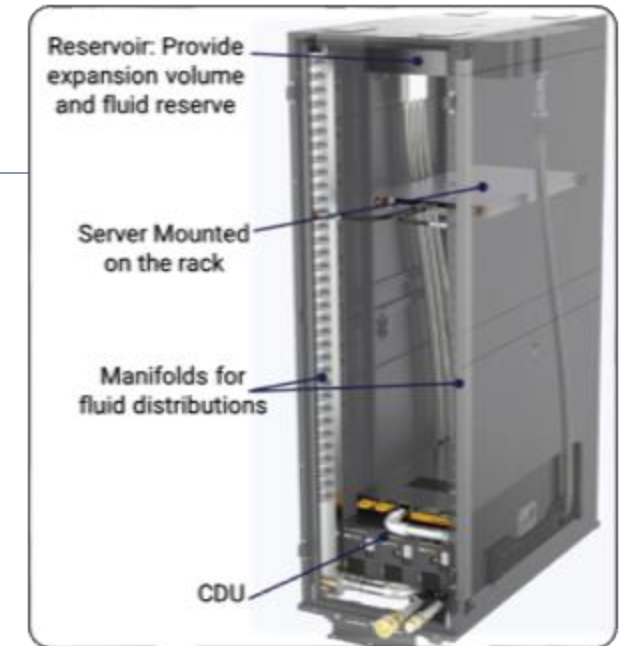
TTV



QTQDs for easy attachment
of the cooling system



Instrumentation
for the fluid side

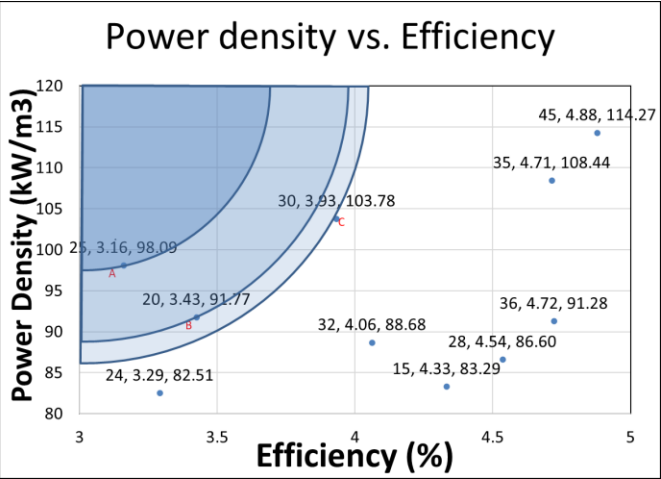


Details of the two-phase CDU:

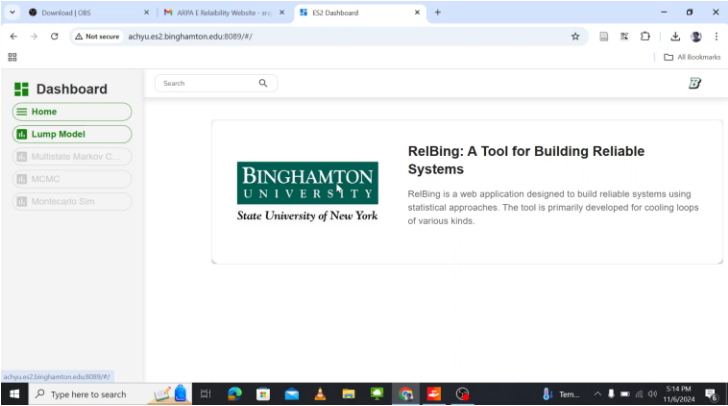
- Coolable Heat: 80 kW
- Refrigerant Type: R1233zd(E)
- Rack Height: 42U
- IPCU Height: 7U
- Multiple redundancies with hot swappable serving
- Industrial condenser with isolated facility water loop uses warm (ASHRAE W45) or chilled water
- Zero risk of water coming in contact with electronics, eliminating risk to GPUs

System Reliability

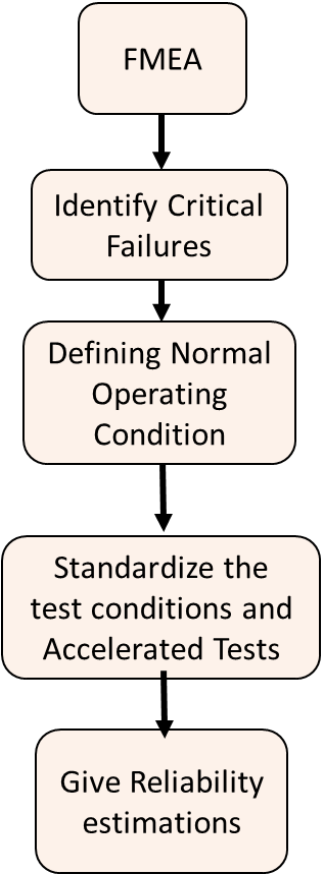
Estimating operational availability



Tools Developed



Reliability approach

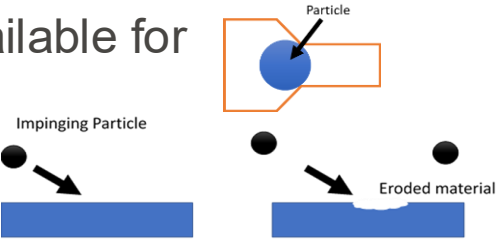


Sl.no	Failure Mode	Criticality	Occurrence	Detection	RPN
1	Delamination	6	3	3	54
2	Cracks in silicon	6	2	5	60
3	Corrosion	5	2	6	60
4	Loose powered fouling	6	5	2	60
5	Erosion	6	3	3	54
6	Leakage	5	2	2	20
7	Non uniform injection	1	4	4	16
8	Crack in Manifold	1	2	2	4

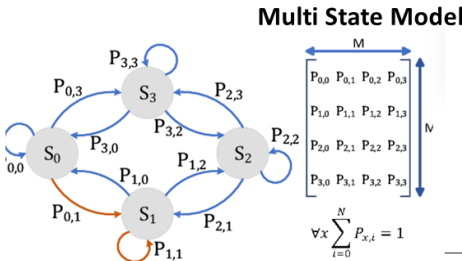
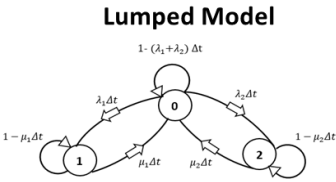
Assessing Reliability: Modeling and testing

Exercising models available for

- Clogging
- Corrosion
- Erosion



State Space Model: Markov Chain



Needs/The Ask: Funding, Partnerships, Advisors

- o *CHiPADD founding team talent (speak to Dr. Schiffres & Dr. Enright today)*
- o *Processor OEM partnership (mitigate technical & business barriers)*
- o *Commercialization funding (scale from lab to field deployed MVP)*

