Embedded Microfluidic Cooling for Nextgen High-Power Server Architectures

Michael Cumbie, HP Inc.

Team Members:
HP Inc.: Paul Benning, Arun Agarwal
NVIDIA: Tom Gray, Tahir Cader

Project Vision

- Our microfluidic Silicon Cold Plate (SiCP) solution will be smaller, lighter, and will deal effectively with die warpage through microbond technology developed at HP: *A microfluidic cooling solution that scales with GPU power and size roadmap*

- This will be achieved by leveraging HP's commercialized 5th generation of inkjet microfluidics to directly couple microfluidics to NVIDIA's GPU chip surface

COOLERCHIPS Kickoff Meeting
October 18 & 19, 2023

Total Project Cost:	$7.77M
Length	36 mo.

Embedded Microfluidic Cooling for Nextgen High Power Server Architectures

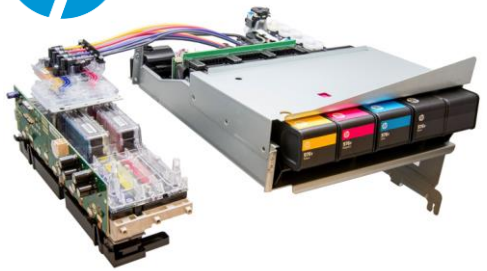
Fed. funding:	$3.25M
Length	36 mo.

Team member	Location	Role in project, core competencies
HP Inc.	Corvallis, Oregon	Silicon microfluidic cooler design and fabrication
NVIDIA	Hillsboro, Oregon; Spokane, Washington	GPU interface and server integration and testing

ARPA-E COOLERCHIPS teaming list brought HP & NVIDIA together in a new way…

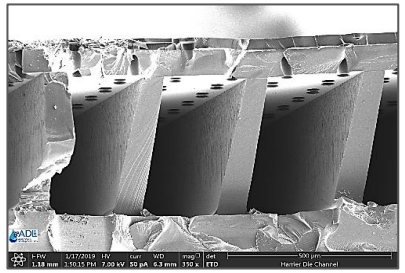
- **Our approach** leverages HP's commercialized 5th generation of inkjet microfluidics platform and relies on first directly coupling silicon microchannels to NVIDIA's GPU chip surface, then by embedding microfluidics deeper into the device as a future step. This is a single-phase cooler that will reject server heat to 40°C and 60% relative humidity external ambient air.
- **Compared to the state-of-the-art** our Silicon Microchannel ColdPlate (SiCP) solution will be smaller, lighter, and will deal effectively with die warpage through microbond technology developed at HP.
- **Our SiCP will be designed as a drop in solution** for single-phase, liquid cooled systems, accelerating time to market by leveraging HP's G5 platform, developed over the last decade for high volume commercial printing applications that also demand an extremely high level of reliability.

Our leverage



Fluid Delivery Systems

- Manifolding macro to micro
- Flow balancing
- Process & assembly



HP's G5 MEMS

- Microchannels on device
- Process / tooling knowledge
- Materials

Nextgen GPU

- State of the art GPU
- Server test systems
- Device integration
- Path to market

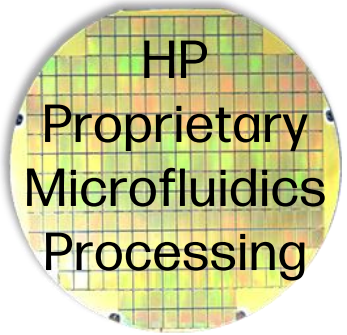
arpa·e
CHANGING WHAT'S POSSIBLE

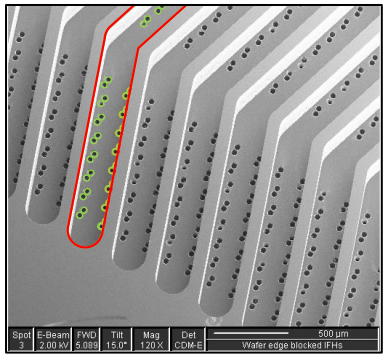
HP G5 Ink Circulation

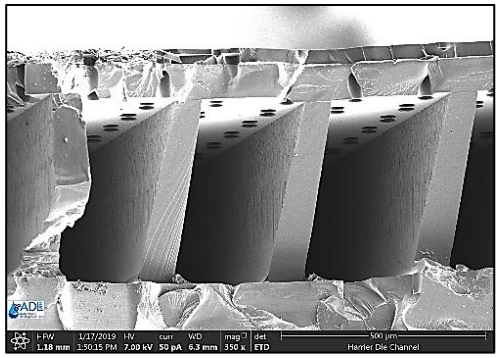
Enables Industrial Performance Leadership

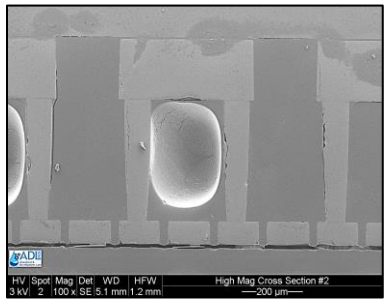
G5 Microfluidics

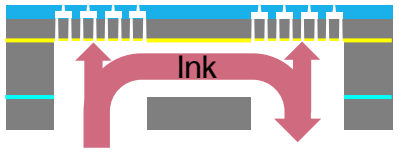
Ink Circulation
Enabled by Through
Silicon Fluid Vias (TSV-F)
And Silicon Microchannels



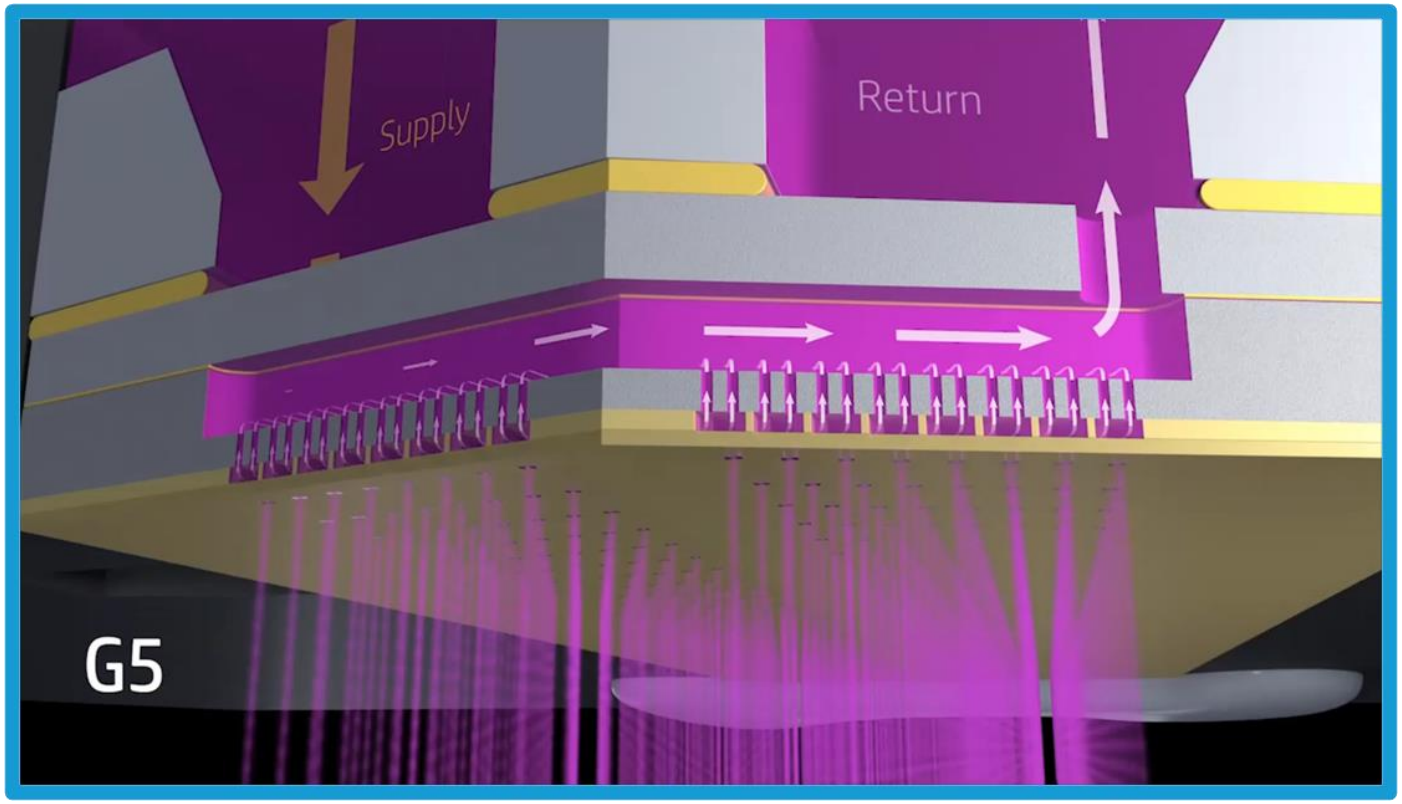








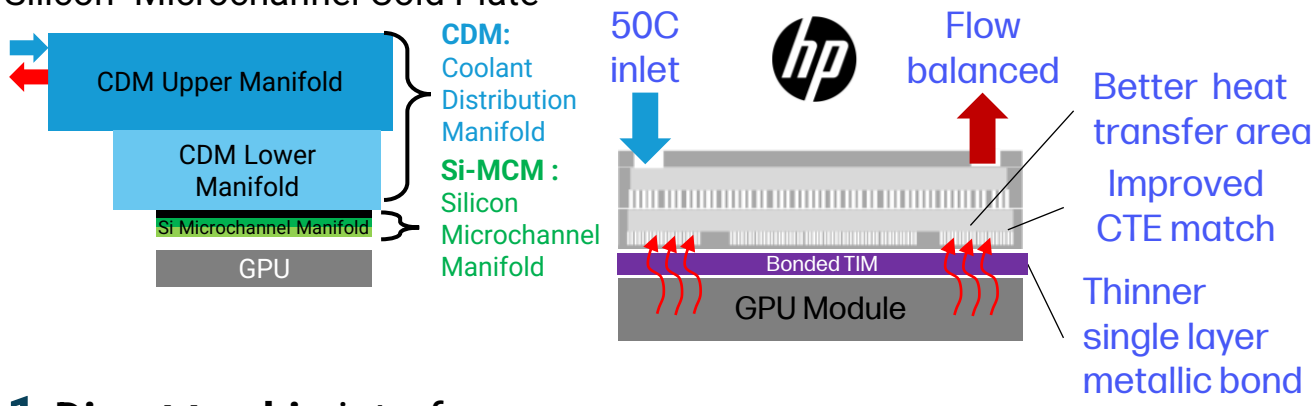
>1 mile of ink channel / wafer



arpa·e
CHANGING WHAT'S POSSIBLE

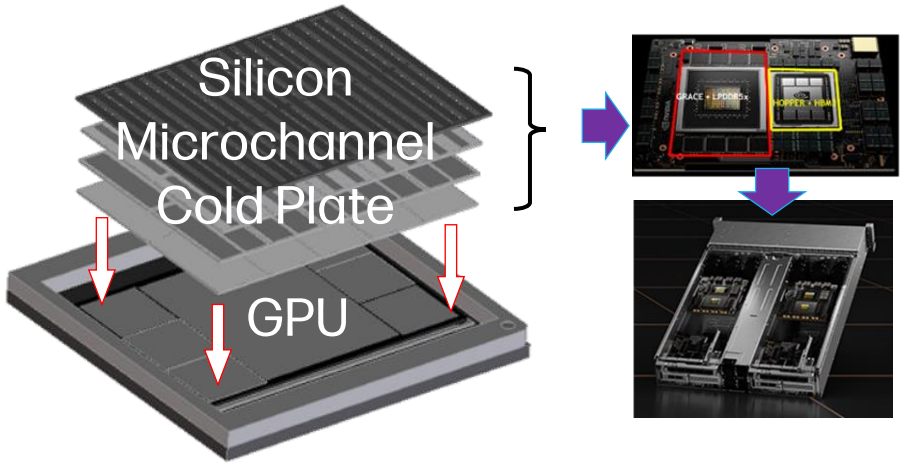
Concept Overview

Silicon Microchannel Cold Plate



1. **Direct to chip** interface
2. **Thinner metallic TIM bond line** by managed warp with proprietary design for < 0.003 K/W resistance
3. **Si microchannels** for improved thermal coupling of ~0.01 K/W chip to coolant resistance and up to 2kW TDP cooling potential at < 3LPM
4. **Flow balanced manifold** for uninform cooling with low pressure drop of < 60kPa @ < 3LPM flow rates
5. **Wafer scale manufacturing** leveraging HP's commercialized G5 multi-layer silicon MEMS processing used in industrial inkjet to speed up time to market, establish reliable baseline.

Silicon Microchannel Cold Plate + Nextgen GPU

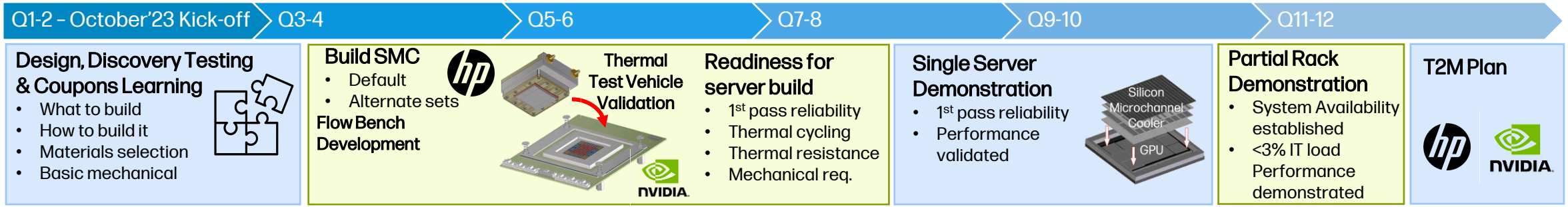


FOA Metrics	Units
Resistance Target	0.01 K/W
Cooling Power % of IT_power	1.27 % on secondary loop
System availability	99.982 %
Chipset	NVIDIA Grace Hopper Next
Chip Power	>1000 W
Power per server	3kW / U, which will scale to over 126kW in a 42U rack
Demonstration power mid project	3 kW 1U server

arpa·e
CHANGING WHAT'S POSSIBLE

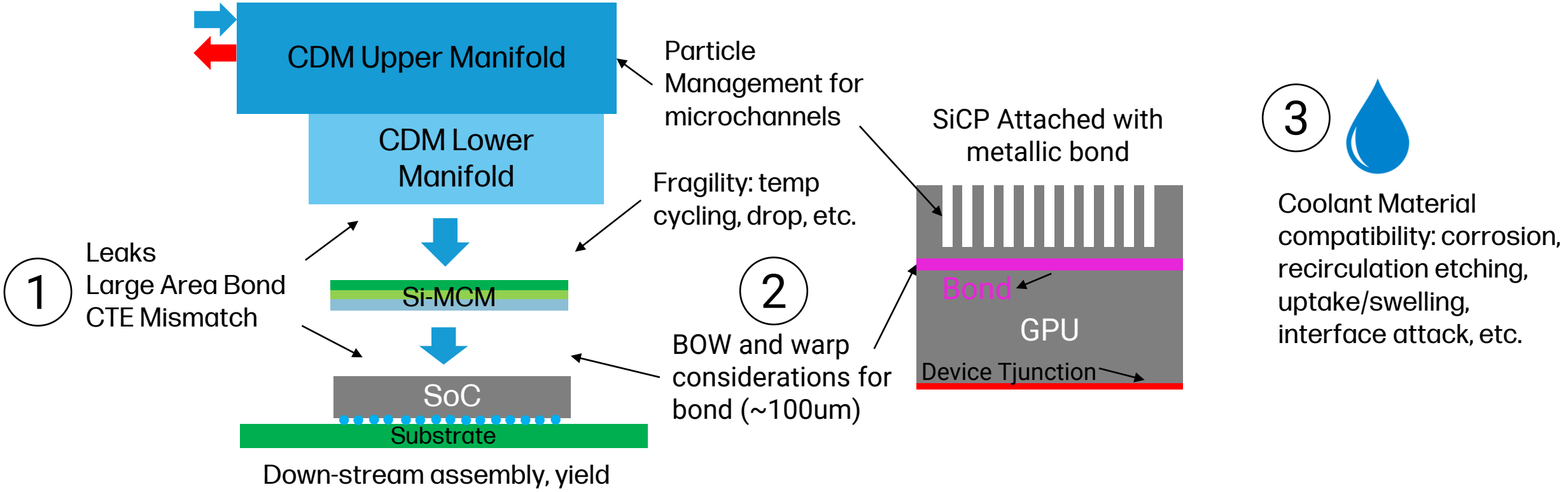
Task Outline & Technical Objectives

- ***Technology Objectives***
  - *Develop advanced silicon microfluidic cold plate (SiCP) technology for Nextgen high-power GPUs and server architectures.*
  - *Demonstrate scalability of the SiCP to high volume manufacturing, and that it can meet reliability and cost targets.*
  - *Demonstrate an SiCP integrated into a single, test server and partial rack.*
- ***Development Activities***
  - *HP: Silicon microchannel cold plate design, modeling, manufacturing, reliability, fluidic/thermal performance testing*
  - *NVIDIA: Server and rack integration, performance testing and reliability analysis; IT System modeling*



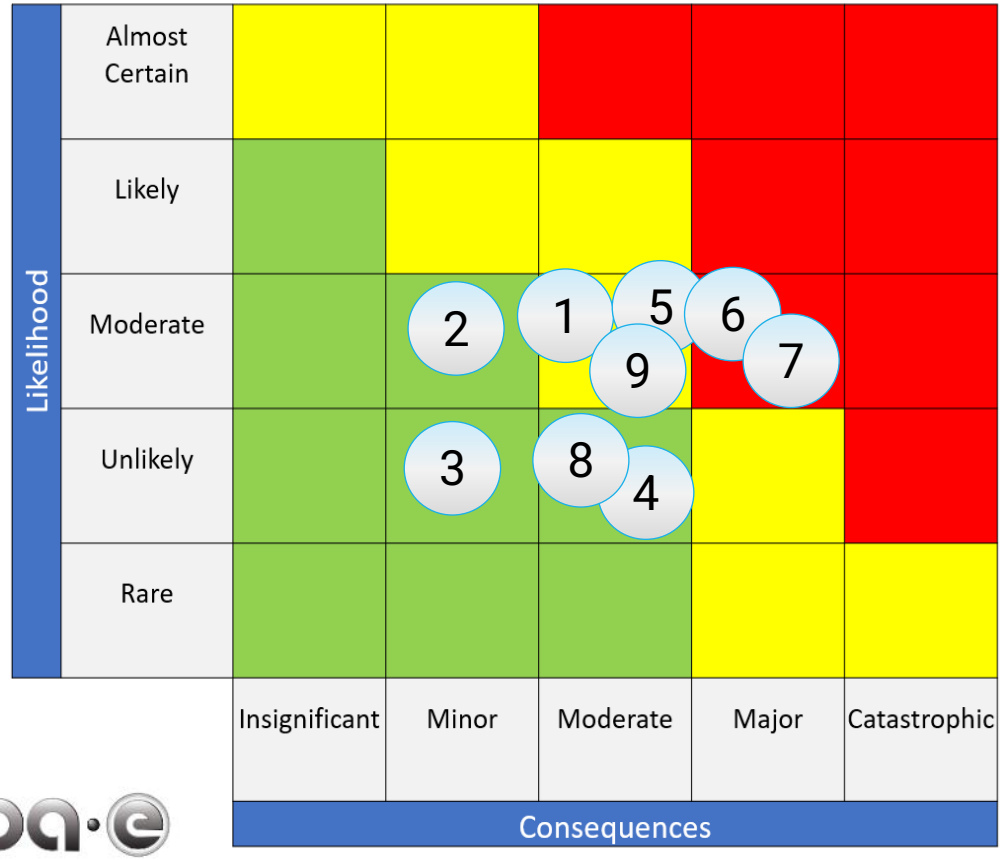
Challenges and Risks

- Key risks are driven primarily by reliability considerations as *silicon microchannel cold plates (SiCP) have never been attempted at scale for chip cooling*



Challenges and Risks

- *We will assess our SMC component failure rate through Fault Tree Analysis (FTA) to inform the System Model*
- *We will assess System Availability using a reliability block diagram (RBD) to determine the appropriate system component's mean-time-between-failure (MTBF) & mean-time-to-repair (MTTR)*



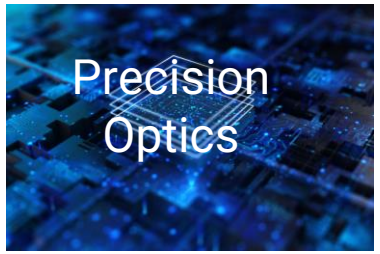
#	Risk
1	Insufficient thermal conductivity of SMC to GPU interface causes > 0.01 K/W
2	Thermal non-uniformity in SMC causes hot spots and < 1.5kW TDP
3	SMC to SMC thermal variation due to poor flow distribution within server causes <1.5kW TDP
4	Pressure drop too high resulting in CDU power requirements exceeding IT budget target of < 3%
5	Structure failure between SMC and GPU
6	Leak in the manifold or fluid connections
7	Materials compatibility with coolant
8	Cost prohibitively high
9	Slow adoption of technology due to cost or performance mismatch

arpa·e
CHANGING WHAT'S POSSIBLE

Technology-to-Market Approach

- Our Silicon Microchannel ColdPlate (SiCP) is a new type of cold plate technology, with the major difference being that it will be directly bonded to the GPU chip as a first step, and embedded deeper into the device package as a future roadmap. **It will be designed as a drop-in cooling solution** replacement for single-phase liquid-cooled cold plates.

- We see this development as the start of **a new technology platform** aimed at producing SiCPs for use in NVIDIA's advanced chip packages or through licensed partners. Once established, we anticipate this platform expanding into areas like microfluidic cooling for high-power electronics, precision optics, and related fields.



- Scaling up fabrication for SiCPs requires steep manufacturing hurdles to be overcome. We will **leverage HP's** extensive experience from its printing business and **5th generation technology to build SiCPs** that meet the fabrication requirements for commercial deployment in NVIDIA's server products initially.

- ARPA-E funding during the proposed project timeline will significantly accelerate the technology time-to-market by helping to offset some of the heavy investment required to deliver a robust, scalable solution; and make delivering a technology like this possible in 3-5 years.

Thank You!

- *We are currently partnering with supply chain in development for materials, coolants, and components to make this vision real. Please let us know of your interest in future collaboration.*

Q & A

https://arpa-e.energy.gov