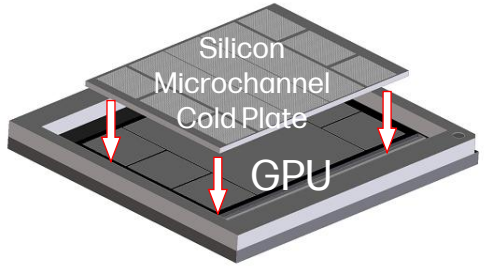
Embedded Microfluidic Cooling for Nextgen High-Power Server Architectures

Michael Cumbie (PI), HP Inc.

Team Members:
HP Inc.: Erik Torniainen, Mark Luckeroth, Paul Benning (T2M)
NVIDIA: Tahir Cader, Matt Slaby, Tom Gray

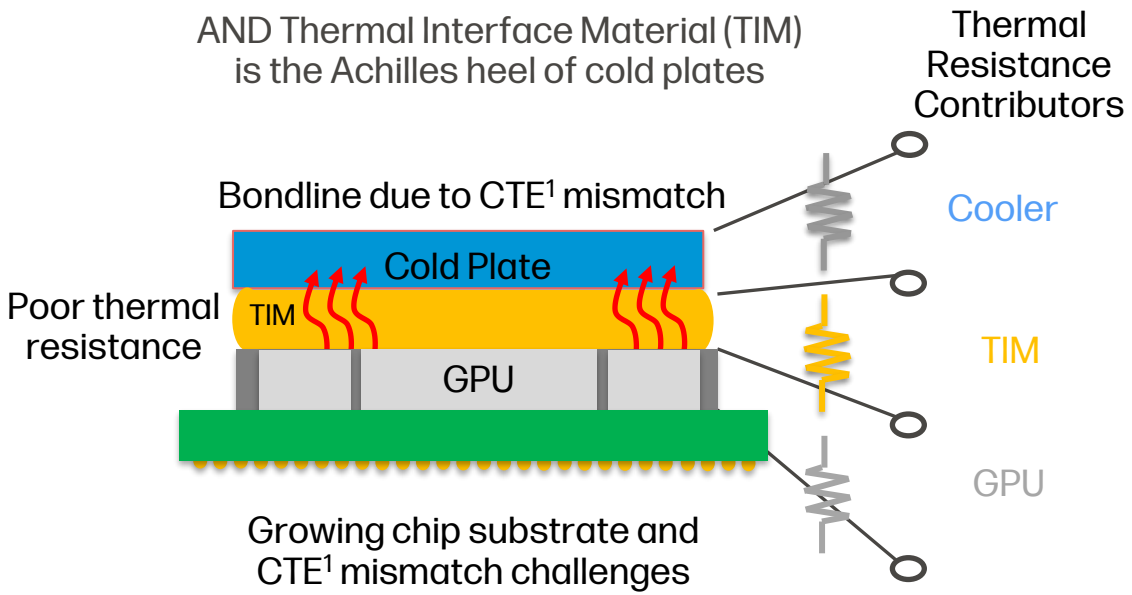
- Our microfluidic Silicon Cold Plate (SiCP) solution will be smaller, lighter, and integrated through microbond technology developed at HP: *A microfluidic cooling solution that scales with GPU power and size roadmap*
- This will be achieved by leveraging HP's 5th generation commercialized inkjet to directly couple microfluidics to NVIDIA's GPU chip surface



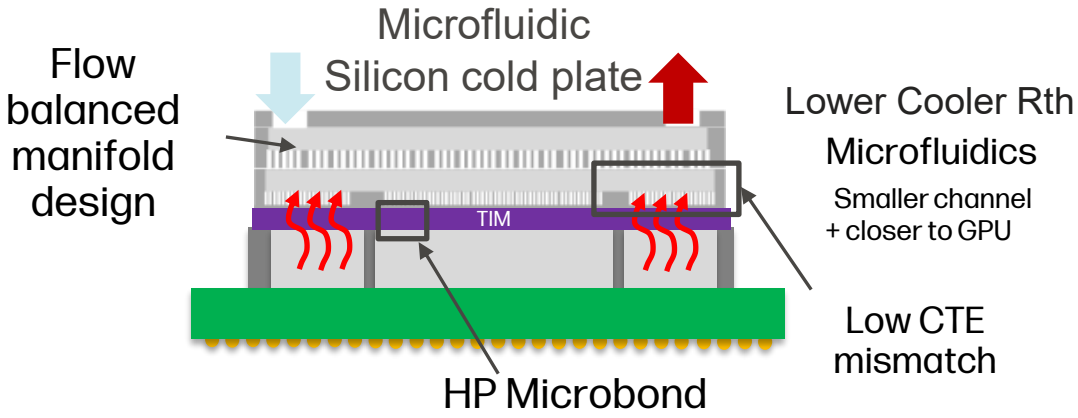
HP silicon microfluidics chip cooling technology approach

Dissipate heat more efficiently and quickly while scaling with the size of the chips

Traditional cold plate manufacturing is more challenged to take advantage of microfluidics



HP Approach: Silicon microfluidics that scale with GPU power and size requirements



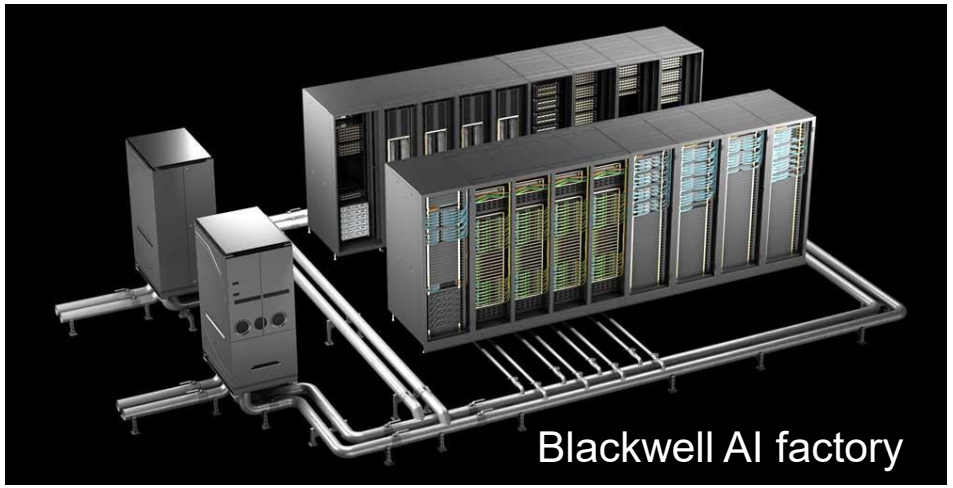
Notes: 1) CTE: Coefficient of Thermal Expansion

Solution / Product / Technology Description And Impact

Compared to the state-of-the-art our microfluidic Silicon Cold Plate (SiCP) solution will be smaller, lighter, and will deal effectively with die warpage through microbond technology developed at HP.

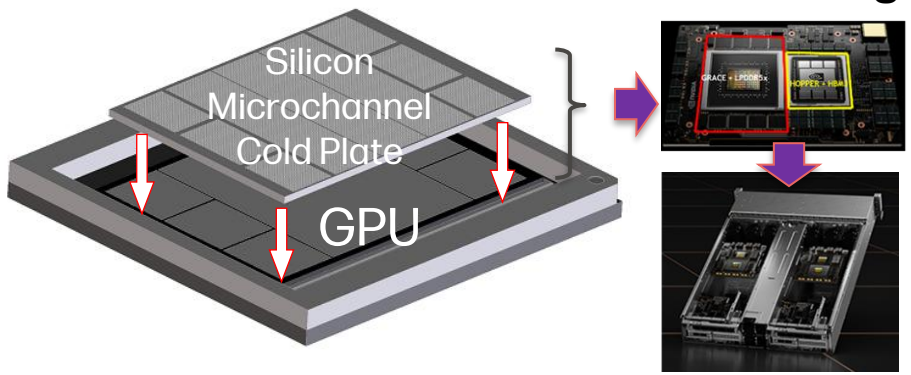
Our SiCPs will be designed as a drop-in solution for single-phase, liquid cooled systems, accelerating time to market by leveraging HP's G5 platform, developed over the last decade for high volume commercial printing applications that demand an extremely high level of reliability.

Drop-in solution for single-phase systems



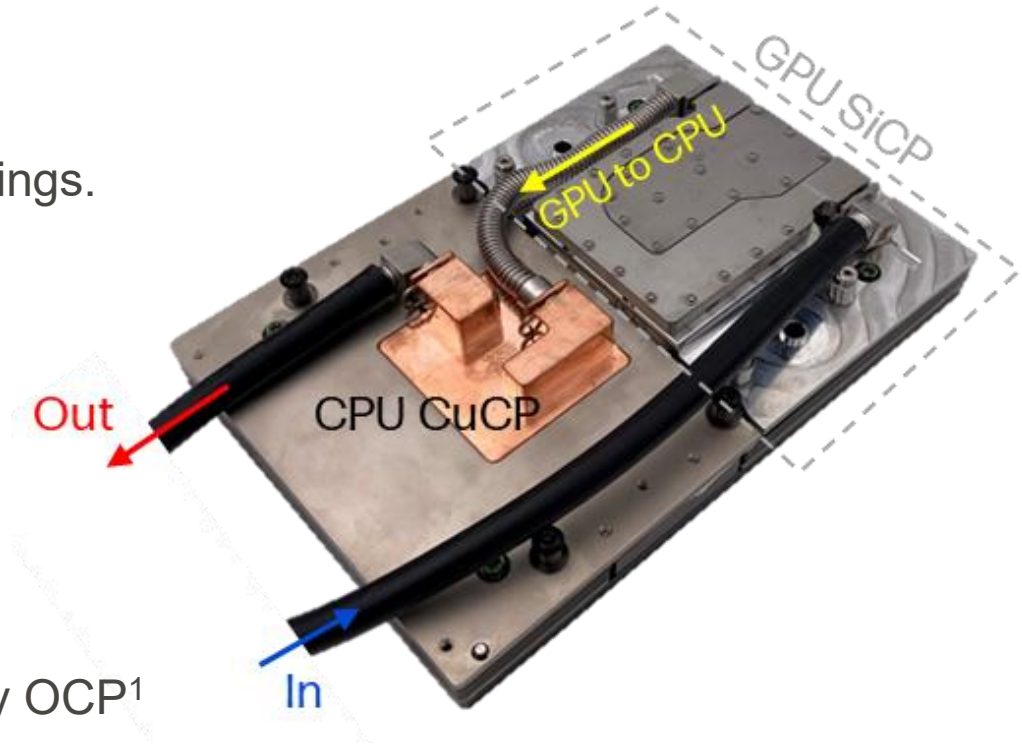
Img src: https://nvidianews.nvidia.com/online-press-kit/computex-2024-news

Silicon Microchannel Cold Plate + Nextgen GPU



Future “Megawatt Rack” Solutions Must Leverage LC Platforms

- Silicon microchannel cold plate fits within the existing volumetric of skived copper cold plates.
  - XY footprint of GPU package
  - Enables Z-height for 1U tray height
- Compatible with existing PG25 and Di-water tray plumbing and fittings.
- Rack design concepts show <30% increase in rack manifold ΔP
  - TCS Loop flow supported by existing CDUs
  - 0.01°C/W, Chip surface to cold plate inlet
- Supported by standard CDUs with 25µm inline filters
  - May require side-stream 5µm filtration, as recommended by OCP[1]
- Demonstrated compatible with both Hopper and Blackwell compute tray designs.



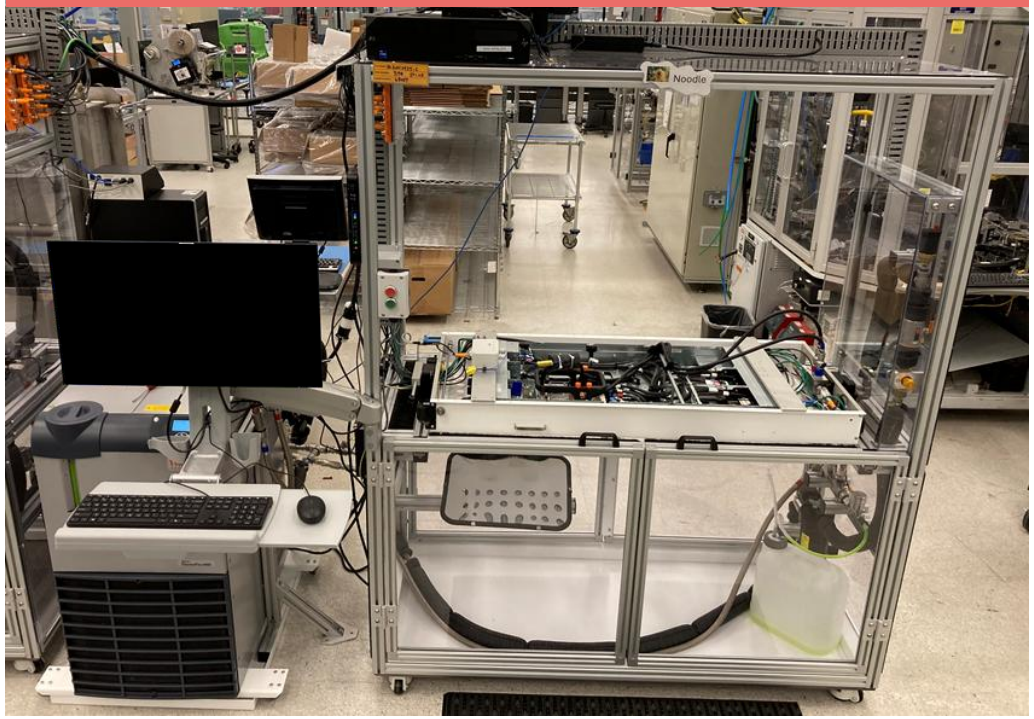
[1] Guidelines for Using Propylene GlycolBased Heat Transfer Fluids in SinglePhase Cold Plate-Based Liquid Cooled Racks

System Integration Results and Midterm Milestone

- First generation HP SiCP Microbond technology has been developed
- Live NVIDIA Hopper server successfully demonstrated and shown repeatable
- ARPA-E Program Target, R-th 0.01°C/W

Hopper Development Vehicle

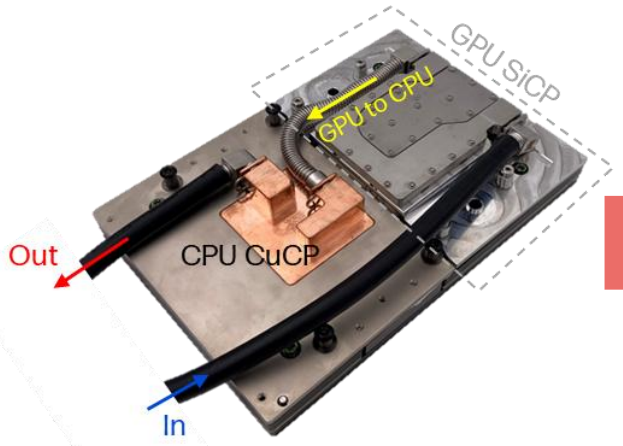
Server Integration Tester



What is Next: Timetable

- Next milestone will demonstrate 0.01°C/W performance on Blackwell.
- The Silicon micro-channel cold plate integrates into the existing GB200 server tray HW and will be tested using a liquid-to-liquid or liquid-to-air CDU.
- Reliability testing with component level results supporting system availability model.

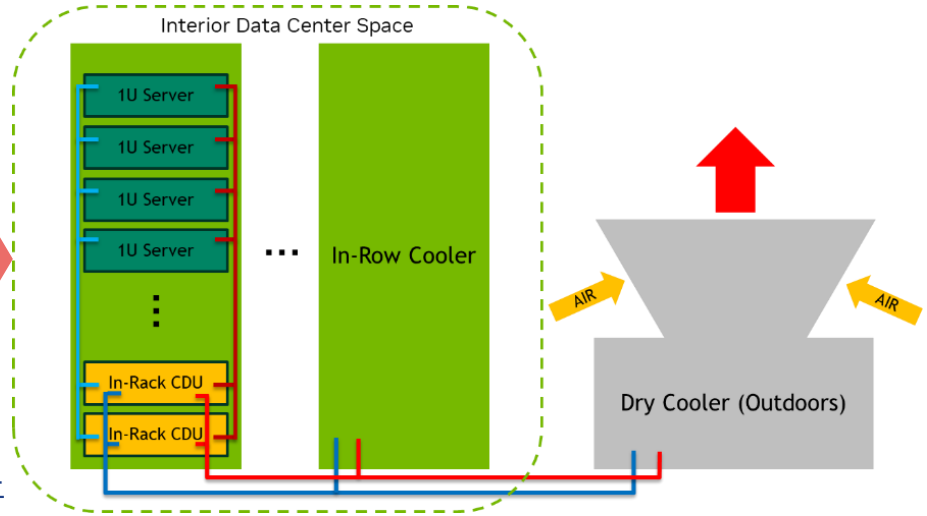
SiCP + Hopper Development Vehicle



Blackwell Live Server Demo <0.01 C/W

https://blogs.nvidia.com/blog/blackwell-platform-water-efficiency-liquid-cooling-data-centers-ai-factories/

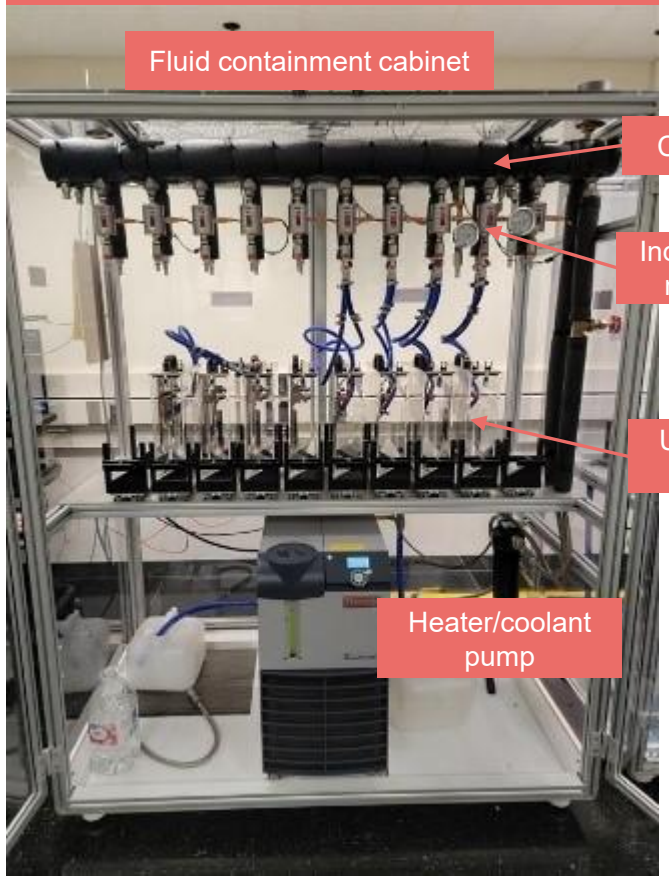
Partial Rack Demo



System Reliability

- Tested using elevated temperature and coolant pressure conditions
- Temperature cycling to validate integration reliability
- Developing standards for new wetted materials compatibility based on OCP validation tests[1] and HP chemistry expertise
- FMEA guided test plan focused on silicon microfluidics

Reliability Integrated Flow Tester: Elevated Temp/Pressure



[1] Guidelines for Using Propylene GlycolBased Heat Transfer Fluids in SinglePhase Cold Plate-Based Liquid Cooled Racks