 

**A Tool for Optimized Cooling of High-Performance Systems**: Cooling the high-performance computing systems in AI datacenters, 5G radio access networks (RAN), and edge compute nodes is a growing challenge. That’s because the multi-chip modules (MCMs) used in these systems, which integrate CPUs, GPUs, and high-bandwidth memory (HBM) in 2.5D/3D architectures, already draw more than 1000 watts, and that will increase going forward. A Nokia Bell Labs team will describe a high-power thermal test vehicle (TTV) they built to assess the efficiency of various novel cooling solutions. At its core is a multi-chip module (MCM) consisting of six emulated logic dies and 12 emulated HBM stacks. These are mounted on a ceramic interposer capable of dissipating more than 2600 W, with localized hotspots reaching 314 W/cm² (and 817 W/cm2 over the heater resistor area). The MCM has 264 platinum-based resistive temperature detectors for high-resolution thermal monitoring. It enables independent power control across 60 heater regions, facilitating the creation of diverse heat maps of discretized regions across the MCM surface. This capability enables the study and optimization of various novel cooling technologies.

In addition, the TTV was integrated into a pumped two-phase refrigerant cooling system, which successfully handled heat fluxes up to 314 W/cm² over a 46 mm2 hotspot on each logic die. By reducing peak chip temperatures as much as 25 °C, it highlights the superior thermal performance of two-phase versus single-phase cooling.

Above:

* **At left** is a photo of the fabricated single-substrate multi-chip module (MCM) with 18 thermal test chips. The six larger chips are the logic dies, while the 12 peripheral chips are high-bandwidth memories.
* **At right** are thermal maps comparing the performance of single- and two-phase flows, represented by the difference in the MCM temperatures (left column) and the ratio of the specific thermal resistances (right column) of single- and two-phase tests conducted at ~ 430 W and ~ 640 W (top and bottom rows, respectively).

**(IP session #40, “*Experimental Demonstration of High-Power Thermal Test Vehicle using Two-Phase Cooling for AI Datacenters, 5G RAN, and Edge Compute Nodes*,” Yang Liu et al, Nokia Bell Labs**)