Thermal Characterization for AI chips

A common requirement for thermal management design efforts of AI chips is the need for a real-time, hardware-oriented, emulation capability. Attempting to use real application chips is difficult to implement for several reasons.

First, waiting until the chips are available means that time-to-market is greatly elongated. Chip customers need several months, or even a year or more, after receipt of the chips to design and implement a production-worthy thermal management solution. This extended time reduces the return on development investment and gives competitors time to get to the market first, thereby establishing creditability, mindshare, and initial customer base.

Second, even if time-to-market considerations are not a high priority, when creating a laboratory setup using application chips(s), the instrumentation for thermal management design and development is not a trivial task, in terms of both time and cost. Getting application chips to dissipate power requires complex circuitry. Additionally, software must be developed to exercise the chip(s) to dissipate sufficient power to provide a heat source for the thermal management design effort.

Third, the instrumentation necessary for a thermal management design effort is not easily implemented using application chips. One example is the physical location of temperature sensors on the chip. Because the power dissipated in the chip is usually not unform across the chip's area, a single temperature sensor located in one portion of the chip may not adequately reflect temperature in another portion of the chip. How can the thermal management take this into account if the different temperatures are not accurately known? The problem is further complicated when the "application chip" is a big CPU chip surrounded by smaller chips, as in a chiplet package, with smaller chips having their own thermal issues to contend with.

Fourth, temperature sensing is not the only issue. Each area of the big main chip and each of the surrounding small chips usually operates at different power density levels, and implementing these power density levels is usually very difficult. On top of this problem, trying to significantly change normal operating power levels to stress the system is even harder to implement. How do you easily modify the development setup and/or software to increase the power dissipation in certain sections (or chips), for example?

These difficulties can be overcome by using a Thermal Test Vehicle (TTV) built around Thermal Test Chips (TTC). The TTV package, which usually matches the physical size and material composition of the application device it is intended to emulate, can contain one or more TTCs each closely matching the size and power dissipation maps of their real-life counterparts in the application device package.

The TTC is an array of Unit Cells (either 1mm square or 2.54mm square) where each cell contains heating resistor(s) and temperature sensing diode(s) (TSD) which can be individually and directly addressed. Power dissipation mapping is accomplished by electrically inter-connecting cells on the package substrate and bringing out the desired connections, and the location-desired TSD connections can be brought out in a similar way to allow for temperature measurements and real-time mapping. Thus, the power density can be controlled, and the temperature can be monitored in real time, down to the Unit Cell level across each of the TTCs in the package.

As a result, the power mapping of the application device can be accurately emulated using the TTV (coupled with actual thermal management solutions, including thermal interface materials

and heat sinks, for example), and the temperature evolution across each chip can be closely monitored, for steady state and transient settings.

For further information about how the TTC/TTV can enable the development of AI device thermal management solutions, please visit www.thermengr.net.