The Importance of Worst-Case Memory Contention Analysis for Heterogeneous SoCs

Lorenzo Carletti, Gianluca Brilli, Alessandro Capotondi, Paolo Valente, and Andrea Marongiu
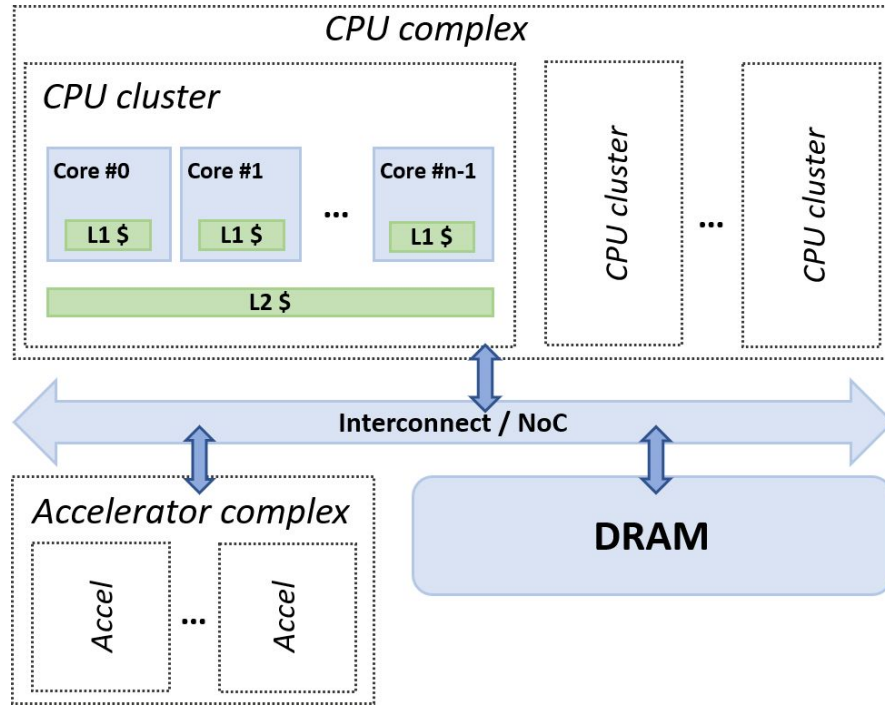
UNIMORE, University of Modena and Reggio Emilia, 41125 Modena, Italy
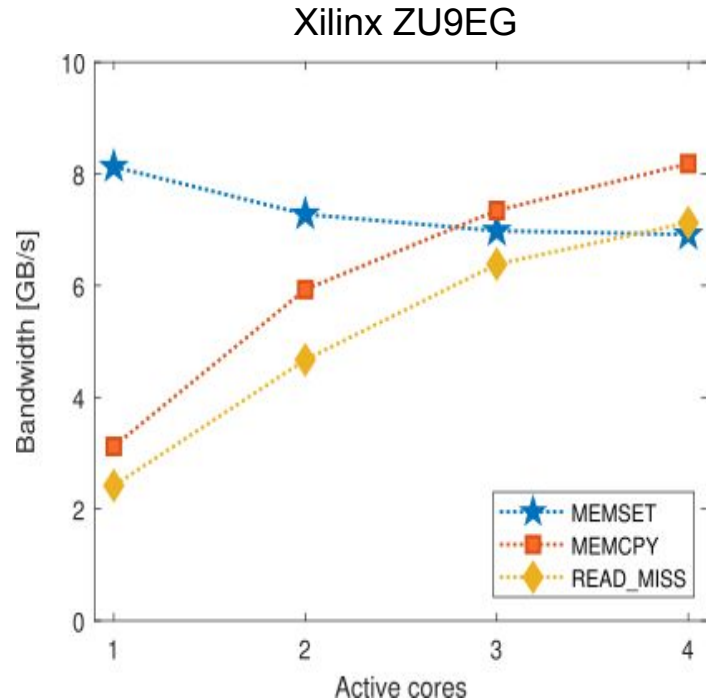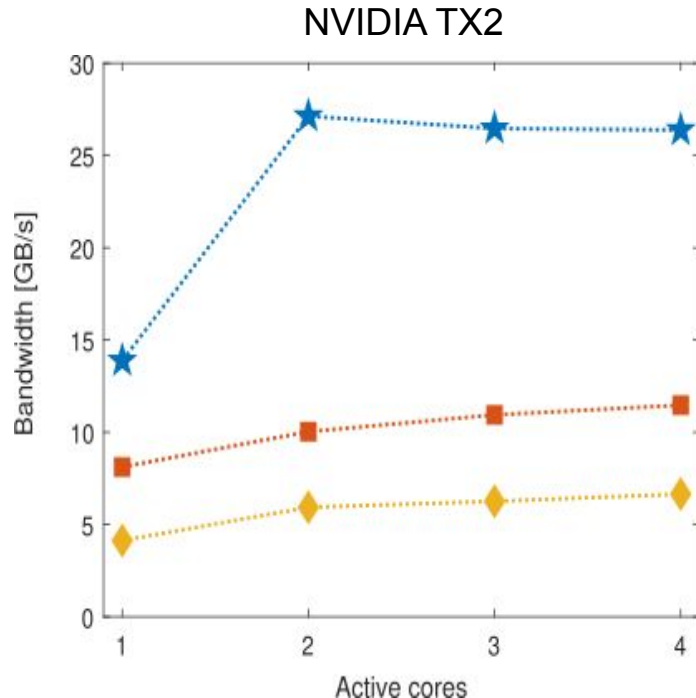
Hardware layout of HeSoCs

Chosen hardware:
- NVIDIA TX2 (GPU)
- Xilinx ZU9EG (FPGA)

A Bandwidth anomaly

Measured bandwidth for various types of traffic using synthetic memory-only benchmarks.
- READ_MISS :Read only
- MEMCPY: Read + Write
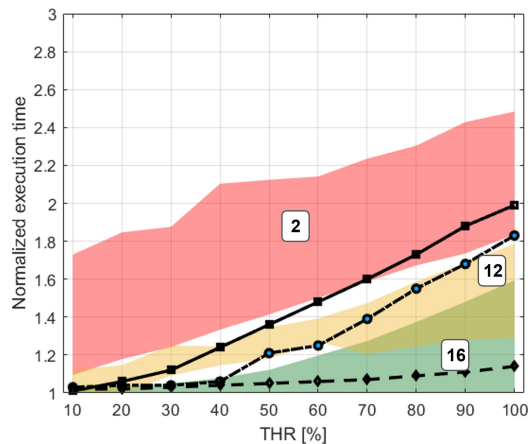- MEMSET: Write only

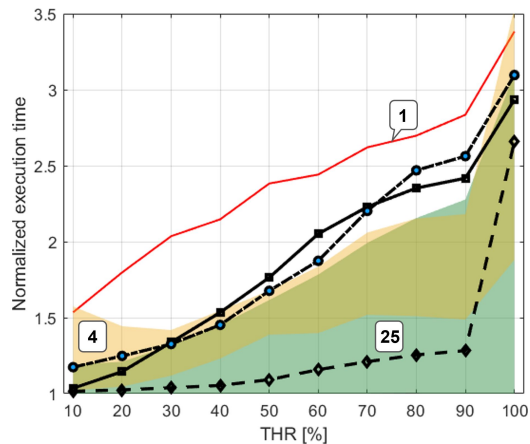NVIDIA TX2

Xilinx ZU9EG

NVIDIA TX2 Interference Analysis

Measured slowdown for Synthetic Benchmarks and Polybench against the Synthetic Benchmarks
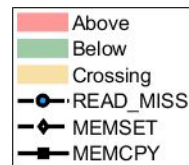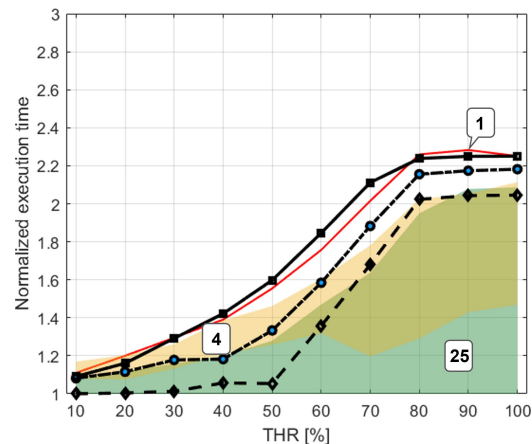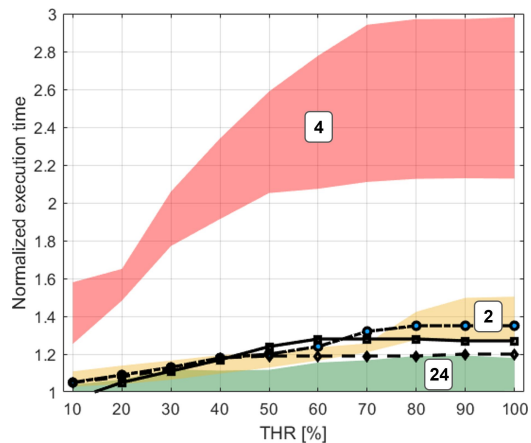- MEMCPY causes the most slowdown
- READ_MISS the most slowed down

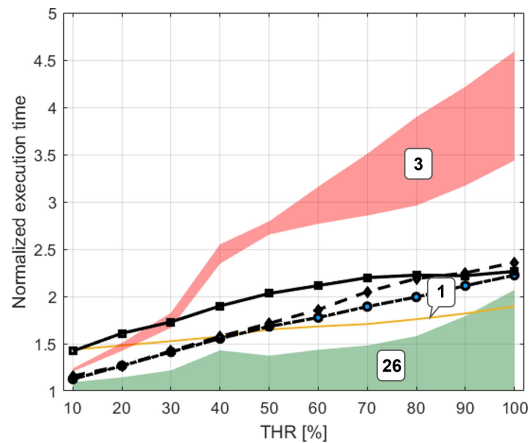# Xilinx ZU9EG Interference Analysis

Measured slowdown for Synthetic Benchmarks and Polybench against the Synthetic Benchmarks
- MEMSET causes the most slowdown
- MEMSET the most slowed down
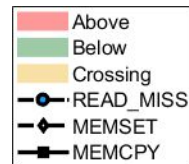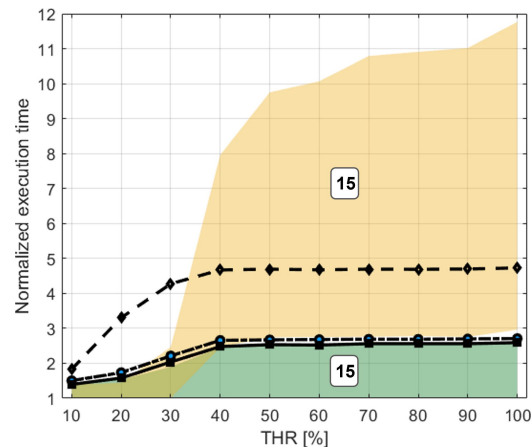
Conclusions

1. The traffic type causing the highest amount of interference is hardware-dependent.
2. The traffic type subject to the highest amount of interference is hardware-dependent.
3. Cache thrashing can cause less memory-intensive benchmarks to suffer more slowdown than the synthetic memory-only benchmarks.

Future work

- Full-on paper on the matter, with a deep dive on the cause for such high latency increase due to cache thrashing.
- Exploration of the effects of using both CPU and FPGA cores to cause interference.
- Long term goal: Study of novel QOS guaranteeing techniques to handle accelerator-based DRAM interference on these kinds of HeSoCs.

Thanks for the attention!