Architectures Session

Dynamic Thermal Management in Heterogeneous Embedded Multiprocessor SoCs

S. Sharifi[1], A. K. Coskun[2], T. S. Rosing[3]
[1]*University of California,San Diego, USA, shervin@ucsd.edu*
[2]*University of California,San Diego, USA, acoskun@cs.ucsd.edu.*
[3]*University of California,San Diego, USA, tajana@ucsd.edu*

Abstract

Multiprocessor system-on-chips (MPSoCs) have become the new design paradigm as they provide higher performance per Watt relative to single core processors. Heterogeneous MPSoCs, consisting of cores with various power/performance characteristics and different architectures, provide a more flexible power-performance trade-off in comparison to homogeneous MPSoCs, especially for the embedded market. To effectively exploit the heterogeneity of the hardware, developing dynamic thermal management methods is necessary to prevent thermal hot spots, as high temperatures increase packaging and cooling costs as well as degrading performance and reliability. The magnitudes of the temperature gradients on the chip are also likely to increase thus accelerating failure mechanisms and introducing timing errors.

In this work, we propose a temperature-aware job scheduling technique integrated with dynamic voltage and frequency scaling. Our technique balances the temperature across the chip while matching the processing capability of the heterogeneous MPSoC with the performance requirements of the current workload. The technique consists of an offline phase and an online (runtime) phase. In the offline phase, first a correspondence is established between the processing capability of MPSoC and the voltage/frequency configurations of cores to create a balanced thermal profile while providing the required performance. Also, the tasks which are expected to run on the system are characterized to determine the core types that best fit each task's characteristics. At runtime, the processing demand of the workload is estimated, and from the look-up table generated in the first phase, the appropriate voltage/frequency settings of cores are selected. Based on the characterization performed offline, each new incoming task is directed to the queue of its preferred core type. Load balancing among these queues prevents cores from getting overloaded. Our experimental results show that we are able to reduce the peak temperature up to 4^{o}C and the energy consumption up to 10%, while meeting the performance requirements, in comparison to an equivalent homogeneous MPSoC.