Adaptive Partitioning Strategies for Loop Parallelism in Heterogeneous Architectures

Angeles Navarro*, Antonio Vilches*, Rafael Asenjo*, Francisco Corbera*
*Universidad de Málaga, Andalucía Tech, Dept. of Computer Architecture, Spain.
{angeles,vilches,asenjo,corbera}@ac.uma.es

Abstract—This paper explores the possibility of efficiently using multicores in conjunction with multiple GPU accelerators under a parallel task programming paradigm. In particular, we address the challenge of extending a parallel_for template to allow its exploitation on heterogeneous systems. The extension is based on a two-stages pipeline engine which is responsible for partitioning and scheduling the chunks into the computational resources. Under this engine, we propose a dynamic scheduling strategy coupled with an adaptive partitioning heuristic that resizes chunks to prevent underutilization and load unbalance of CPUs and GPUs. In this paper we introduce the adaptive partitioning heuristic which is derived from an analytical model that minimizes the load unbalance while maximizes the throughput in the system. Using two benchmarks we evaluate the overhead introduced by our template extensions finding that it is negligible. We also evaluate the efficiency of our adaptive partitioning strategies and compared them with related work.

REFERENCES