Dynamic Workload Distribution for Solidification Modeling on Multi- and Manycore Platforms

Roman Wyrzykowski, Kamil Halbiniak, Lukasz Szustak, Adam Kulawik
Institute of Computer and Information Sciences,
Czestochowa University of Technology, Czestochowa, Poland
{roman, khalbiniak, lszustak, akulawik}@icis.pcz.pl

Providing an appropriate load balancing across computing resources plays a significant role in optimizing the overall performance of computations. It allows maximizing the performance of applications by keeping the idle time as low as possible. Numerous mechanisms for ensuring efficient workload distribution and data partitioning have been developed. However, the load balancing problem is not solved completely. It is evident especially in the scientific applications where the workloads are unpredictable or change during the execution. Such codes require developing dynamic strategies of load balancing that adapt the workload distribution based on changes monitored during the application execution

In our previous work, an application dedicated to the numerical modeling of alloy solidification was used as a testbed in the suitability assessment of various programming environments for porting a real-life scientific application to hybrid shared-memory platforms with multicore CPUs and manycore Intel MIC accelerators. Two different cases were considered: with the static and dynamic intensity of computations. In the first case, the workload of computing resources is constant during the application execution. The solidification application withthe dynamic intensity assumes that calculations are performed for a carefully selected group of nodes. The usage of suitable selection criterion allows reducing the amount of computations. However, the consequence is a significant workload imbalance, since the selection criterion is calculated after partitioning of the grid nodes between computing resources.

In this work, we focus on developing a method for adapting the solidification application with the dynamic intensity of computations to multi-/manycore architectures with shared memory. The main goal here is to ensure an efficient data partitioning and workload distribution across computing resources. To achievethis goal, we propose algorithms which allow us dynamically manage computing resources during the application execution. These algorithms are based on predicting the amount of computations performed in successive time step of simulation. The performance of developed approach is examined for two platforms:(i) SMP consisting of two Intel Xeon Platinum 8180 CPUs (Intel Xeon Scalable Processor architecture) and (ii) single Intel Xeon Phi 7250F processor with the KNL ar-

chitecture. The achieved performance results show that the proposed algorithms allow reducing the execution time of simulation significantly.

Keywords: multicore, manycore, Intel Xeon Scalable processor, KNL processor, numerical modeling of solidification, OpenMP, load balancing, workload prediction.