FACILITATING EMERGING APPLICATIONS ON MANY-CORE PROCESSORS

Da Li

Dr. Michela Becchi, Dissertation Supervisor

ABSTRACT

Over the last decade, many-core Graphics Processing Units (GPUs) have been widely used to accelerate a variety of applications. Meanwhile, Intel has released its Xeon Phi Coprocessor, which is equipped with more than fifty x86 cores, each supporting four hardware threads. Despite their widespread use, many-core processors are still considered difficult to program.

Due to their massive computational powers, many-core processors have been successfully used to parallelize a wide variety of dense matrices and vectors based applications. These extensively investigated problems are mainly from linear algebra, stencil computations, image processing and so on. However, many established and emerging problems have not yet been fully explored. It has been shown that these emerging applications do exhibit certain degree of static and runtime parallelism, but are relatively hard to parallelize.

My research focuses on addressing important issues related to the deployment of emerging applications on many-core processors. In particular, we proposed efficient GPU implementations for large-scale pairwise sequence alignment. We also targeted graph- or tree-based applications and proposed: (1) unifying programming interfaces for many-core processors (2) runtime support for efficient execution on irregular datasets and (3) compiler support for efficient mapping of applications onto hardware. Finally, we conducted a comprehensive study of performance, memory footprint and power consumption on various platforms and extended existing central processing units (CPU) only or graphic processing units (GPU) only CNNs learning methods to CPU-GPU cooperative ways. We also implemented a virtual memory and integrated into Caffe to facilitate training large CNN models with limited GPU memory.