Towards a multifrontal QR factorization for heterogeneous architectures over runtime systems

Florent Lopez

Université de Toulouse, INPT(ENSEEIHT)-IRIT, France

Abstract

During the last decade, computer architectures for high performance computing have considerably evolved toward heterogeneous systems equipped with different types of computational units and a higher number of cores per chips. An example of popular heterogeneous architectures widely adopted in the high performance computing domain are GPU-based systems.

In the work presented in this talk we study the exploitation of heterogeneous architectures for sparse direct solvers. To achieve this goal we consider the use of a runtime system proved to be extremely efficient and robust for the implementation of dense linear algebra algorithms. Sparse linear algebra algorithms, however, are commonly characterized by complicated data access patterns, computational tasks with extremely variable granularity and complex dependencies. Therefore, a substantial research effort is necessary to design and implement sparse direct methods capable of achieving high efficiency as well as reliable and accurate solutions on emerging architectures.

Through this work, we propose original methods to efficiently exploit direct methods on top of a runtime system and performance results show the potential of this approach on multicore architecture. Furthermore, we present preliminary performance result on hybrid CPU-GPU systems in order to extend our approach to heterogeneous architectures.