

Fast and reliable linear system solutions on new parallel architectures

Marc Baboulin
Université Paris-Sud/Inria

Abstract

Recent years have seen an increase in peak "local" speed through parallelism in terms of multicore processors and GPU accelerators. At the same time, the cost of communication between memory hierarchies and/or between processors have become a major bottleneck for most linear algebra algorithms.

In this presentation we explain how hybrid multicore+GPU systems can be used efficiently to enhance performance of linear algebra libraries. We illustrate this approach by considering hybrid factorizations where we split the computation over a multicore and a graphic processor and where the amount of communication is significantly reduced.

Next we describe a randomized algorithm that accelerates factorization of general or symmetric indefinite systems on multicore or hybrid multicore+GPU systems. Randomization prevents the communication overhead due to pivoting, is computationally inexpensive and requires very little storage. The resulting solvers outperform existing routines while providing us with a satisfying accuracy.