Abstract

Sparse systems of linear equations are computational bottlenecks in applications ranging from science to optimization. For many problems, including Mechanical Computer Aided Engineering (MCAE), iterative methods are unreliable and sparse matrix factorization is performed. Multifrontal sparse matrix factorization is often preferred and, by representing the sparse problem as a tree of dense systems, maps well to modern memory hierarchies. This allows effective use of BLAS-3 dense matrix arithmetic kernels. Graphics processing units (GPUs) are architected differently than general-purpose hosts and have an order-of-magnitude more single-precision floating point processing power. This paper explores the hypothesis that GPUs can accelerate the speed of a multifrontal linear solver, even when only processing a small number of the largest frontal matrices. We show that GPUs can more than double the throughput of the sparse matrix factorization. This in turn promises to offer a very cost-effective speedup to many problems in disciplines such as MCAE.

Keywords: Computational solid mechanics and materials and application performance