Extreme-speed scalable direct sparse solvers for heterogeneous CPU/GPU/FPGA supercomputing – an enhancement to the LAPACKrc library

Juan G. Gonzalez and Rafael Camilo Nunez

Abstract:

At the core of DOE’s scientific priorities is a universal need for increased computational power in HPC infrastructures. An estimated 70% of HPC cycles worldwide are spent solving linear algebra kernels at both large and small scales. Linear algebra is at the core and constitutes the primary bottleneck of important DOE research problems in fusion energy, nuclear accelerator modeling, circuit simulation, and weather modeling, among other research challenges. As part of an ambitious research program funded by DOE, NASA, and the Department of Defense, Accelogic is spearheading the development of LAPACKrc, a groundbreaking family of FPGA-based linear algebra solvers able to achieve speedups larger than 100x with a single chip (“rc” stands for “reconfigurable computing,” a radically new computing paradigm that is changing the way high performance computing is done today).

In this poster, we present our latest advances in scalable direct sparse solvers for heterogeneous CPU/GPU/FPGA supercomputing, as part of our LAPACKrc effort targeting peta- and exascale performance. The emphasis is on a first generation large-scale sparse Cholesky factorization prototype that has reached 120x acceleration for certain large matrices.