

Ginkgo's load-balancing COO SpMV on NVIDIA and AMD GPU architectures

Tuesday, 5 November 2019 11:20 (30 minutes)

Efficiently processing unbalanced and irregular matrices on manycore architectures is a challenging problem. With the load-balancing Sparse Matrix Vector Multiplication (SpMV) based on the coordinate format (COO), we have designed an SpMV kernel that provides attractive performance across a wide range of matrices. In this contribution, we present the load-balancing COO SpMV kernel, elaborate on architecture-specific tuning knobs, and evaluate the performance and efficiency across different GPU architectures from AMD and NVIDIA. In this evaluation, we do not exclusively focus on runtime performance, but also consider metrics like bandwidth utilization, energy consumption, and performance-per-\$. We also discuss the Ginkgo library in which the presented SpMV functionality is available. Ginkgo is a next-generation sparse linear algebra library able to run on multi- and manycore architectures. The library design is guided by combining ecosystem extensibility with heavy, architecture-specific kernel optimization. The library design is guided by combining ecosystem extensibility with heavy, architecture-specific kernel optimization. The software development cycle ensures production-quality code by featuring unit testing, automated configuration and installation, Doxygen code documentation, as well as a continuous integration and continuous benchmarking framework. Ginkgo is an open source effort licensed under BSD 3-clause and ships with the latest version of the xSDK package (v.0.5.0).

Primary authors: TSAI, YU-HSIANG (Karlsruhe Institute of Technology); ANZT, Hartwig (Karlsruhe Institute of Technology); COJEAN, Terry (Karlsruhe Institute of Technology); GRÜTZMACHER, Thomas (Karlsruhe Institute of Technology); NAYAK, Pratik (Karlsruhe Institute of Technology); RIBIZEL, Tobias (Karlsruhe Institute of Technology)

Presenters: TSAI, YU-HSIANG (Karlsruhe Institute of Technology); COJEAN, Terry (Karlsruhe Institute of Technology)

Session Classification: Day I