Skysong Innovations
Advancing the Arizona State University Knowledge Enterprise

1475 N. Scottsdale Road, Suite 200
Scottsdale, AZ 85287-3538
Phone: 480 884 1996 Fax: 480 884 1984

Case ID:M22-050P

Published: 1/3/2023

High Performance Computing Framework for Accelerating Sparse Cholesky Factorization on FPGAs

Inventors

Erfan Bank Tavakoli
Fengbo Ren
Michael Riera
Masudul Quraishi

Contact

Physical Sciences Team

-Solving large symmetric sparse linear systems using sparse Cholesky factorization plays a pivotal role in many scientific computing and high-performance computing (HPC) applications. The existing computational solutions to sparse Cholesky factorization based on CPUs and GPUs suffer from very limited performance for to two primary reasons. First, sparse Cholesky factorization algorithms are recursive and have complex data dependencies on intermediate results from previous iterations. An algorithm-tailored buffering scheme for efficiently storing the intermediate results must be employed. Unfortunately, the architecture of CPUs and GPUs cannot be adapted to efficiently implement such an algorithm-tailored buffering scheme. Consequently, CPU- and GPU-based solutions suffer from poor cache locality and often require frequent off-chip memory access for computing sparse Cholesky factorization. Second, sparse Cholesky factorization algorithms involve complex operations that are often computed using approximation algorithms that are iterative and have strong loop-carried data dependency. Unfortunately, the architectures of CPUs and GPUs lack the capability to exploit the temporal/pipeline parallelism that is critical to resolving such loop-carried data dependency, which results in long loop initiation intervals. Also, CPU- and GPU-based sparse Cholesky factorization solutions suffer from high energy consumption due to high runtime and power consumption.

FPGAs are being deployed as an alternative solution to accelerating sparse Cholesky factorization for HPC applications. FPGAs can address the performance concerns of CPU- and GPU-based solutions. FPGAs can implement an algorithm-tailored buffering scheme, minimize loop initiation intervals, and consume less energy than CPUs and GPUs. However, a framework has yet to be developed which can take advantage of FPGAs to accelerate sparse Cholesky factorization.

Researchers at Arizona State University have developed a framework for accelerating sparse Cholesky factorization on FPGAs. The proposed framework includes a deeply pipelined and scalable FPGA kernel that accelerates supernodal multifrontal Cholesky factorization algorithm and a scheduling algorithm for efficient assignment of computational nodes for elimination tree-based multifrontal methods. Framework can eliminate the need for off-chip memory access for storing intermediate results.

Related publication: FSCHOL: An OpenCL-based HPC Framework for Accelerating Sparse Cholesky Factorization on FPGAs

Potential Applications:

- High performance computing (HPC) applications including:
  - Linear least squares
  - Non-linear optimizations
  - Monte Carlo simulations
  - Kalman filters
  - Matrix inversion

Benefits and Advantages:

- Minimizes on-chip memory requirements for intermediate results
- ASU's framework exhibits on average one order of magnitude higher performance and lower energy consumption compared to the state-of-the-art implementations of sparse Cholesky factorization on CPU, GPU, and other FPGA work