Taking graphics cards beyond gaming

January 10 2017

The graphics cards found in powerful gaming computers are now capable of solving computationally intensive mathematical problems common in science and engineering applications, thanks to a new solver developed by researchers from the KAUST Extreme Computing Research Center.

"One of the most common problems in scientific and engineering computing is solving systems of multiple simultaneous equations involving thousands to millions of variables," said David Keyes, KAUST Professor of Applied Mathematics and Computational Science, who also led the research team. "This type of problem comes up in statistics, optimization, electrostatics, chemistry, mechanics of solid bodies on Earth and gravitational interactions among celestial bodies in space."

In typical applications, solving such problems is often the main computational cost. Thus, acceleration of the solver has the potential to considerably impact both the execution time and the energy consumption required to solve the problem.

"Graphics processing units (GPUs) are very energy efficient compared with standard high-performance processors because they eliminate a lot of the hardware required for standard processors to execute general-purpose code," explained Keyes. "However, GPUs are new enough that their supporting software remains immature. With the expertise of Ali Charara, a Ph.D. student in the Center who spent several months as an intern at NVIDIA in California, we have been able to identify many things that we can either innovate or improve upon, such as redesigning a

common solver."

The key to making a more efficient solver is maximizing the trade-off between the number of processors and the memory available to temporarily store the computational data. Memory remains expensive, so finding a way to execute more computation using less memory is critical to solving the problem of computational cost.

"Charara designed a solver scheme that operates directly on data 'in place' without making an extra copy," explained Hatem Ltaief, a Senior Research Scientist from the project's team. "This means a system twice as large can be stored in the same amount of memory."

Charara achieved this by converting operations that are often carried out by progressing sequentially over columns in the data into a series of tasks on small, computationally efficient, rectangular and triangular blocks recursively carved out of the matrix. This performed specifically on columns in the matrix of values derived from the set of simultaneous equations. This redesigned triangular matrix-matrix multiplication implementation achieves up to eightfold acceleration compared to that of existing implementations.

"Now, every user of an NVIDIA GPU has a faster solver for a common task in scientific and engineering computing at their disposal," said Keyes.

The solver is due to be integrated into the next scientific software library for NVIDIA GPUs.

  More information: Charara, A., Ltaief, H.,& Keyes, D. Redesigning triangular dense matrix computations on GPUs. In European Conference on Parallel Processing, pp. 477-489. Springer International Publishing (2016). DOI: 10.1007/978-3-319-43659-3_35