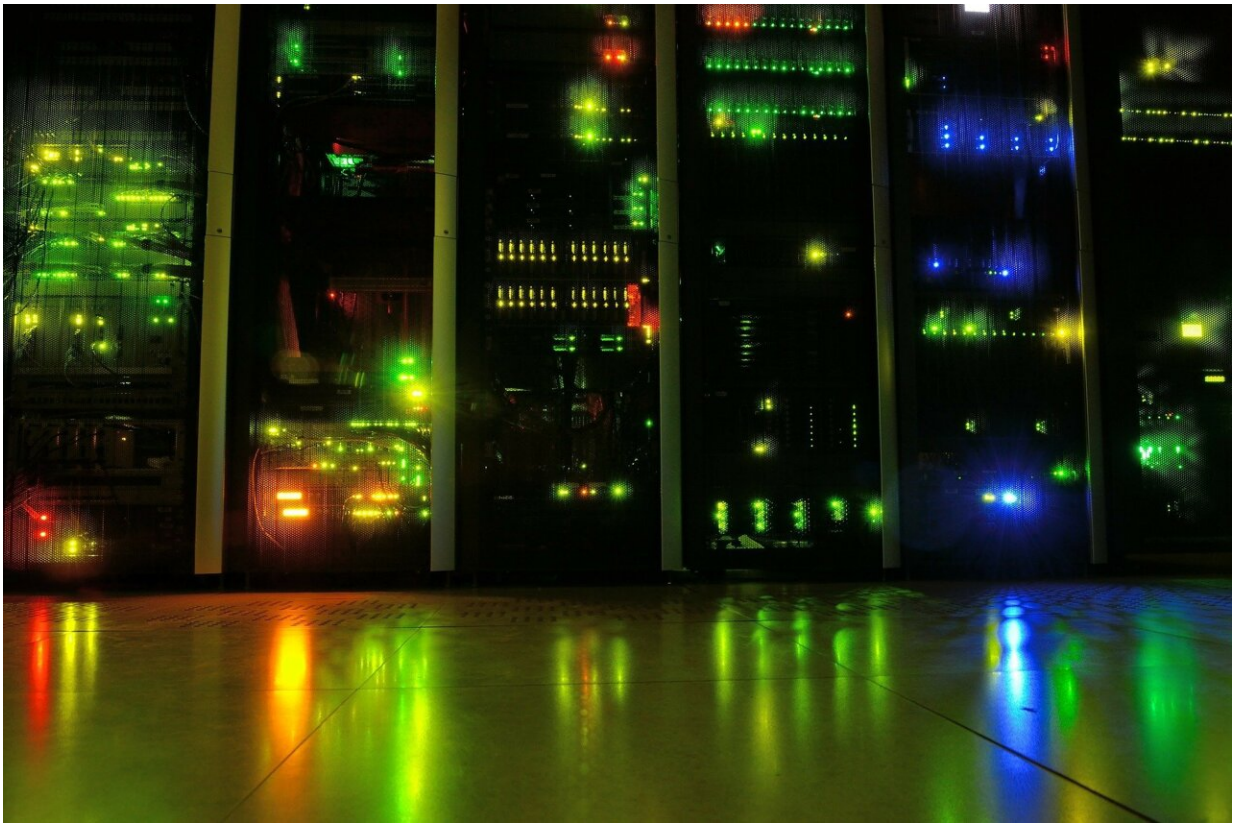
Open-source GPU technology for supercomputers

April 30 2021



Credit: Pixabay/CC0 Public Domain

Researchers from the HSE International Laboratory for Supercomputer Atomistic Modeling and Multi-scale Analysis, JIHT RAS and MIPT have compared the performance of popular molecular modeling

programs on GPU accelerators produced by AMD and Nvidia. In a paper published by the *International Journal of High Performance Computing Applications*, the scholars ported LAMMPS on the new open-source GPU technology, AMD HIP, for the first time.

The scholars thoroughly analyzed the performance of three molecular modeling programs—LAMMPS, Gromacs and OpenMM—on GPU accelerators Nvidia and AMD with comparable peak parameters. For the tests, they used the model of ApoA1 (Apolipoprotein A1)—apolipoprotein in blood plasma, the main carrier protein of 'good cholesterol." They found that the performance of research calculations is influenced not only by hardware parameters, but also by software environment. It turned out that ineffective performance of AMD drivers in complicated scenarios of parallel launch of computing kernels can lead to considerable delays. Open-source solutions still have their disadvantages.

In the recently published paper, the researchers were the first to port LAMMPS on a new open-source GPU technology, AMD HIP. This developing technology looks very promising since it helps effectively use one code both on Nvidia accelerators and on new GPUs by AMD. The developed LAMMPS modification has been published as an open source and is available in the official repository: users from all over the world can use it to accelerate their calculations.

"We thoroughly analyzed and compared the GPU accelerator memory sub-systems of Nvidia Volta and AMD Vega20 architectures. I found a difference in the logics of parallel launch of GPU kernels and demonstrated it by visualizing the program profiles. Both the memory bandwidth and the latencies of different levels of GPU memory hierarchy as well as the effective parallel execution of GPU kernels—all these aspects have a major impact on the real performance of GPU programs," said Vsevolod Nikolskiy, HSE University doctoral student

and one of the paper's authors.

The paper's authors argue that participation in the technological race of the contemporary microelectronics giants demonstrates an obvious trend toward greater variety of GPU acceleration technologies.

"On the one hand, this fact is positive for end users, since it stimulates competition, growing effectiveness and the decreasing cost of supercomputers. On the other hand, it will be even more difficult to develop effective programs due to the need to consider the availability of several different types of GPU architectures and programming technologies," said Vladimir Stegailov, HSE University professor. "Even supporting program portability for ordinary processors on different architectures (x86, Arm, POWER) is often complicated. Portability of programs between different GPU platforms is a much more complicated issue. The open-source paradigm eliminates many barriers and helps the developers of big and complicated supercomputer software."

In 2020, the market for graphic accelerators experienced a growing deficit. The popular areas of their use are well-known: cryptocurrency mining and machine learning tasks. Meanwhile, scientific research also requires GPU accelerators for mathematical modeling of new materials and biological molecules.

"Creating powerful supercomputers and developing fast and effective programs is how tools are prepared for solving the most complex global challenges, such as the COVID-19 pandemic. Computation tools for molecular modeling are used globally today to search for ways to fight the virus," said Nikolay Kondratyuk, researcher at HSE University and one of the paper's authors.

The most important programs for mathematical modeling are developed by international teams and scholars from dozens of institutions.

Development is carried out within the open-source paradigm and under free licenses. The competition of two contemporary microelectronics giants, Nvidia and AMD, has led to the emergence of a new open-source infrastructure for GPU accelerators' programming, AMD ROCm. The open-source character of this platform gives hope for maximum portability of codes developed with its use, to supercomputers of various types. Such AMD strategy is different from Nvidia's approach, whose CUDA technology is a closed standard.

It did not take long to see the response from the academic community. Projects of the largest new supercomputers based on AMD GPU accelerators are close to completion. The Lumi in Finland, with 0.5 exaFLOPS of performance (which is similar to performance of 1,500,000 laptops!) is quickly being built. This year, a more powerful supercomputer, Frontier, is expected in the U.S. (1.5 exaFLOPS), and in 2023—an even more powerful El Capitan (2 exaFLOPS) is expected.

More information: Nikolay Kondratyuk et al, GPU-accelerated molecular dynamics: State-of-art software performance and porting from Nvidia CUDA to AMD HIP, *The International Journal of High Performance Computing Applications* (2021). DOI: 10.1177/10943420211008288

Provided by National Research University Higher School of Economics