

Соотношение между энергопотреблением и производительностью GPU на примере алгоритмов молекулярной динамики

В.С. Вечер^{1,2}, В.П. Никольский^{1,3}, В.В. Стегайлов¹

¹Объединенный институт высоких температур РАН

²Московский физико-технический институт (государственный университет)

³Национальный исследовательский университет Высшая школа экономики

Использование графических ускорителей для научных расчетов, в частности – для задач молекулярной динамики, уже давно является обыденным делом. Увеличение размеров исследуемых систем до миллионов атомов, а также постепенное усложнение используемых моделей приводит к соответственному увеличению требований к размерам доступной памяти и вычислительной мощности. В то же время, постепенная эволюция GPU превратила их в компактные высокопроизводительные устройства по доступной цене.

Однако, кроме стандартных для сегодняшнего дня тенденций на увеличение производительности и доступной памяти вычислительных устройств, все более актуальным в свете перспективных платформ экзафлопсного класса становится тенденция понижения затрат энергии на расчет.

Хотя Nvidia постоянно заявляет о существенном повышении производительности и энергоэффективности их устройств с каждым новым поколением графических ускорителей, это не отменяет необходимости непосредственной проверки эффективности GPU на реальных вычислительных задачах. К сожалению, применяемые в молекулярной динамике алгоритмы несколько отличаются от принятых в отрасли стандартных вычислительных тестов (вроде High Performance Linpack). Данный факт приводит к идее применения используемого на практике молекулярно-динамического кода в качестве альтернативы стандартным бенчмаркам, что позволяет соотносить повышение пиковой производительности вычислительных устройств с получаемым на практике ускорением.

В нашем выступлении мы проведем параллели между различными поколениями графических ускорителей от Nvidia и версиями ускоренных на GPU алгоритмов в контексте молекулярно-динамического пакета LAMMPS и его модулей USER-CUDA, KOKKOS и GPU.

Для сравнения вычислительной эффективности и энергозатрат на расчет мы использовали гибридные миникомпьютеры Jetson TK1 (32 бит) и TX1 (64 бит) с интегрированными графическими ядрами архитектуры Kepler и Maxwell. Применяя бенчмарк Empirical Roofline Tool, а также внешние источники SmartPower с интегрированными ваттметрами, мы измеряли затраченную на расчет энергию и соотносили ее с полученной в тесте ERT пиковой производительностью. Используя LAMMPS со стандартной моделью жидкости Леннард-Джонса в качестве бенчмарка, мы соотносили полученное ускорение и энергозатраты на разных поколениях графических ядер и на разных версиях гибридных алгоритмов LAMMPS с измеренной пиковой производительностью и энергоэффективностью доступного оборудования.

Также мы обсудим технологию динамического изменения частот в процессе расчета и ее эффективность в контексте повышения энергоэффективности. Мы покажем существование оптимальных режимов для расчета.

Данная работа поддерживается грантом РНФ14-50-00124.

Литература

1. V. V. Stegailov, N. D. Orekhov, and G. S. Smirnov. HPC Hardware Efficiency for Quantum and Classical Molecular Dynamics // Proceedings of Parallel Computing Technologies: 13th International Conference, PaCT 2015, Petrozavodsk, Russia, 2015. P. 469-473.
2. Nikolsky V., Stegailov V., Vecher S. Efficiency of Tegra K1 and X1 Systems-on-Chip for Classical Molecular Dynamics // Proceedings of the 14th International Conference on High Performance Computing & Simulation (HPCS-2016), Innsbruck, Austria. 2016. P. 682-689. doi:10.1109/HPCSim.2016.7568401