

Проблемы использования технологии NVIDIA CUDA при реализации вычислительных алгоритмов

ЕФРЕМОВ АЛЕКСАНДР АЛЕКСАНДРОВИЧ

Институт вычислительного моделирования СО РАН (Красноярск), Россия
e-mail: aedampir@gmail.com

КАРЕПОВА ЕВГЕНИЯ ДМИТРИЕВНА

Институт вычислительного моделирования СО РАН (Красноярск), Россия

ВЯТКИН АЛЕКСАНДР ВЛАДИМИРОВИЧ

Институт вычислительного моделирования СО РАН (Красноярск), Россия

Работа посвящена рассмотрению проблем использования технологии NVIDIA CUDA при реализации ресурсоемких вычислительных алгоритмов. В теории, использование графических процессорных ускорителей (ГПУ) общего назначения для ускорения расчетов позволяет достичь значительного прироста производительности вычислений. Тем не менее, особенности аппаратной архитектуры графических процессоров нередко служат причиной обратного явления. При попытке использования ГПУ для научных и практических расчетов вычислитель зачастую сталкивается с минимальным или даже отрицательным ускорением. Более того, при детальном сравнении наблюдаются расхождения полученных результатов расчетов на ГПУ и ЦПУ, не связанные с принципами распараллеливания или ошибками реализации.

В данной работе, авторами предпринята попытка описать и структурировать причины возникновения вышеописанных явлений, а также предложить варианты их возможного решения. Несмотря на то, что авторы столкнулись с указанными проблемами при реализации параллельного алгоритма решения уравнения неразрывности модифицированным методом траекторий, схожие проблемы могут возникнуть при использовании ГПУ для ускорения расчетов в гидро- и газодинамике, а также других задач, которые связаны с активным использованием чисел с плавающей точкой.

Среди особенных моментов использования ГПУ для расчетов общего назначения, авторы выделяют следующие:

1. Влияние аппаратной архитектуры ГПУ на точность вычислений.
2. Вычисления на ГПУ в условиях аппаратных ограничений по памяти.

В работе приведены примеры расхождения результатов расчетов на ГПУ и ЦПУ, а также способы решения проблемы нехватки регистровой памяти на ГПУ. Описаны общие подходы, которые используют авторы при адаптации последовательного вычислительного алгоритма к использованию графических ускорителей с поддержкой вычислений общего назначения при помощи технологии NVIDIA CUDA.

Работа выполнена при финансовой поддержке РФФИ (заявка № 14-01-31203).

Список литературы

1. Вяткин А.В., Ефремов А.А., Каропова Е.Д., Шайдуров В.В. Параллельная реализация модифицированного метода траекторий для уравнения неразрывности // Седьмая Сибирская конференция по параллельным и высокопроизводительным вычислениям / Под ред. проф. А.В. Старченко. – Томск: Изд-во Том. ун-та, 2014. – ISBN 978-5-7511-2063-4. – С. 5-13.
2. Whitehead N., Fit-Florea A. “Precision & Performance: Floating Point and IEEE 754 Compliance for NVIDIA GPUs”, NVIDIA Technical White Paper, 2011.
3. Sanders J., Kandrot E. CUDA by Example: An Introduction to General-Purpose GPU Programming. Addison-Wesley, 2010. – 232 p.