

Использование технологии OpenCL для моделирования сейсмических процессов сеточно-характеристическим методом

А. М. Иванов, Н. И. Хохлов

Московский физико-технический институт (государственный университет)
Россия, 141701, г. Долгопрудный, Институтский пер., 9

В работе рассматривается применение технологии OpenCL для моделирования сейсмических процессов. Решается двумерная задача распространения сейсмических волн. Для этого находится численное решение волнового уравнения упругости. Решение данного уравнения находится с использованием сеточно-характеристического метода. Он применим в случае, если система уравнений гиперболическая, а значит, результаты данной работы применимы и к другим задачам, которые сводятся к решению гиперболических систем. Было произведено сравнение производительности численного метода на центральном процессоре (CPU) и на графическом процессоре (GPU). На GPU тесты с использованием технологии OpenCL сравнивались с теми же тестами с использованием технологии CUDA. Производительность алгоритма при использовании технологии OpenCL отличается незначительно от производительности при использовании технологии CUDA, но OpenCL позволяет писать программы для более широкого ряда графических устройств. Вычисления производились как с двойной точностью, так и с одинарной. По сравнению с одним ядром центрального процессора Intel Xeon E5-2697 удалось получить ускорение в 44 раза при использовании графической карты GeForce GTX 780 Ti в вычислениях с одинарной точностью. При этом производительность была равна 460 ГФлопсам. Для вычислений с двойной точностью наибольшее ускорение – в 55 раз на GPU Tesla K80. Полученная производительность – 138 ГФлопс. На GPU серии Tesla была достигнута большая производительность при расчетах с двойной точностью, на GPU Geforce – при расчетах с одинарной точностью. Однако GPU серии Radeon показали средний результат по сравнению с другими картами. Кроме производительности в ГФлопсах и ускорения по сравнению с центральным процессором, для каждого графического устройства была определена доля от пиковой производительности, теоретически достижимой на данном GPU. Использование технологии OpenCL для реализации сеточно-характеристического метода позволило задействовать до 32% от пиковой производительности графических карт, что намного больше, чем удается достичь большинству реализаций расчетных алгоритмов.

Ключевые слова: математическое моделирование, сеточно-характеристический метод, гиперболические уравнения, параллельное программирование, GPU, OpenCL

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 15-07-01931 А.

© 2016 Иванов Андрей Михайлович, Хохлов Николай Игоревич

1. Введение

В настоящее время существует множество задач геофизики и сейсмологии, достаточно точное решение которых может быть получено только методами математического моделирования. В большинстве случаев для численного решения этих задач используется большое количество вычислительных ресурсов [Petrov, Khokhlov, 2014]. С развитием технологий параллельного программирования удается все более эффективно задействовать возможности графических устройств (GPU). Большинство работ в области численного моделирования рассматривает применение технологии CUDA [Castro et al., 2013; Khanna, 2013].

В данной работе рассматривается применение технологии OpenCL, причем эффективность ее использования сравнивается с технологией CUDA, а также с расчетами на CPU. Рассматривается сеточно-характеристический метод [Muratov, 2014] с явной схемой, который относительно просто поддается распараллеливанию.

2. Математическая модель

Для описания поведения среды использовалась модель идеального изотропного линейно-упругого материала. Распространение сейсмических волн в упругой среде описывается следующей системой уравнений:

$$\begin{cases} \rho \dot{\vec{v}} = \nabla \cdot T, \\ T = \lambda(\nabla \cdot \vec{v})I + \mu(\nabla \otimes \vec{v} + \vec{v} \otimes \nabla). \end{cases}$$

Здесь ρ – плотность среды, λ, μ – параметры Ламе, T – тензор напряжений, \vec{v} – вектор скорости. Данная система дифференциальных уравнений в частных производных описывает состояние элементарного объема упругого материала в приближении малых деформаций. В двумерном случае она имеет вид:

$$\begin{cases} \rho \frac{\partial v_x}{\partial t} = \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y}, & \rho \frac{\partial v_y}{\partial t} = \frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y}, \\ \frac{\partial \sigma_{xx}}{\partial t} = (\lambda + 2\mu) \frac{\partial v_x}{\partial x} + \lambda \frac{\partial v_y}{\partial y}, & \frac{\partial \sigma_{yy}}{\partial t} = \lambda \frac{\partial v_x}{\partial x} + (\lambda + 2\mu) \frac{\partial v_y}{\partial y}, & \frac{\partial \sigma_{xy}}{\partial t} = \mu \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} \right). \end{cases} \quad (1)$$

Здесь v_x и v_y – горизонтальная и вертикальная составляющие скорости частиц среды; $\sigma_{xx}, \sigma_{yy}, \sigma_{xy}$ – компоненты тензора напряжения. Систему (1) можно переписать в матричной форме:

$$\frac{\partial u_p}{\partial t} + A_{pq} \frac{\partial u_q}{\partial x} + B_{pq} \frac{\partial u_q}{\partial y} = 0, \quad (2)$$

где u – вектор из 5 независимых переменных $u = (\sigma_{xx}, \sigma_{yy}, \sigma_{xy}, v_x, v_y)^T$. Явный вид матриц A_{pq}, B_{pq} представлен в [LeVeque, 2002]. Здесь и далее подразумевается суммирование по повторяющимся индексам. Собственные значения матриц A_{pq} и B_{pq} таковы:

$$s_1 = -c_p, s_2 = -c_s, s_3 = 0, s_4 = c_s, s_5 = c_p,$$

где c_p и c_s – скорости распространения продольных и поперечных волн в среде.

3. Численный метод

Применяя покоординатное расщепление, можно свести задачу построения разностной схемы для системы уравнений (2), к задаче построения разностной схемы для системы вида:

$$\frac{\partial u_p}{\partial t} + A_{pq} \frac{\partial u_q}{\partial x} = 0. \quad (3)$$

Для гиперболической системы уравнений (3) матрицу A можно представить в виде $A = \Lambda R^{-1}$, где Λ – диагональная матрица, элементы которой — собственные значения матрицы A , а R – матрица, состоящая из правых собственных векторов матрицы A . Введём новые переменные: $w = R^{-1}u$ (так называемые инварианты Римана). Тогда система уравнений (3) сведётся к системе из 5 независимых скалярных уравнений переноса.

Далее для каждого скалярного уравнения переноса выписывается разностная схема численного решения, находится значение на следующем шаге w^{n+1} из значения на предыдущем шаге w^n . После этого восстанавливается решение: $u^{n+1} = R w^{n+1}$.

3. Тесты производительности

Графическое устройство	Количество ядер	Частота, МГц	ГФлопс (SP)	SP:DP	ГФлопс (DP)
GeForce GT 640	384	900	691	24	29
GeForce GTX 480	480	1401	1345	8	168
GeForce GTX 680	1536	1006	3090	24	129
GeForce GTX 760	1152	980	2258	24	94
GeForce GTX 780	2304	863	3977	24	166
GeForce GTX 780 Ti	2880	876	5046	24	210
GeForce GTX 980	2048	1126	4612	32	144
Tesla M2070	448	1150	1030	2	515
Tesla K40m	2880	745	4291	3	1430
Tesla K80	2496	562	2806	1.5	1870
Radeon HD 7950	1792	800	2867	4	717
Radeon R9 290	2560	947	4849	8	606

Таблица 1. Характеристики графических устройств для тестовых расчетов.

Для проведения тестов производительности рассматривалась двумерная тестовая задача. Вычисления производились на сетке с размерами 4096×4096 . Выполнялось 6500 шагов по времени. В следующей таблице представлены устройства, на которых производился запуск тестовой задачи.

В таблице 1 изображены характеристики графических процессоров, на которых производились тестовые расчеты. Сокращения SP и DP использованы для обозначения одинарной и двойной точности расчетов. SP:DP – соотношение между числом операций одинарной и двойной точности за такт.

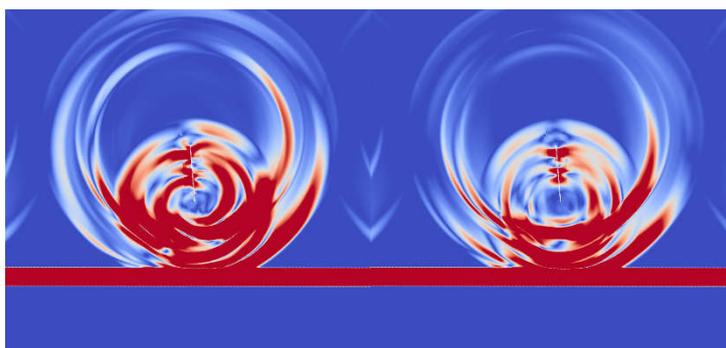


Рис. 1. Пример решаемой задачи [Golubev, 2015] – изображения сейсмических волн. Акустическое решение слева и решение, основанное на физически корректных граничных условиях справа.

Тестовые расчеты были произведены на центральном процессоре и на графических ускорителях, производительность алгоритма с использованием технологии OpenCL сравнивалась с производительностью того же алгоритма написанного с использованием технологии CUDA для GPU фирмы NVIDIA.

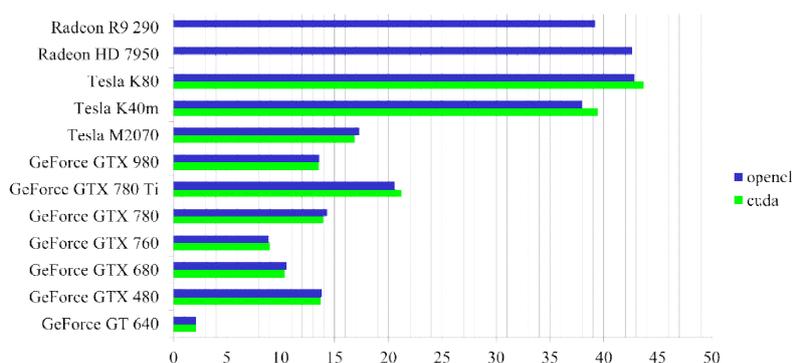


Рис. 2. Ускорение по сравнению с CPU Intel Xeon E5-2697

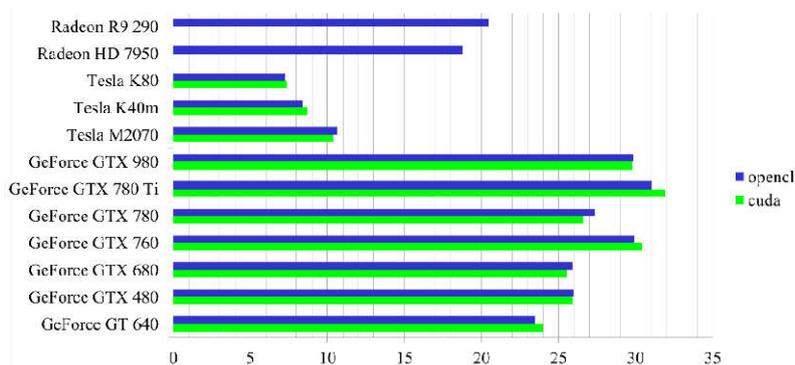


Рис. 3. Процент от пиковой производительности

Было определено ускорение сравнению с одним ядром центрального процессора (рис. 2). Из характеристик графических устройств, представленных в таблице 1 возможно определить процент от пиковой производительности, достижимой на каждом из GPU (рис. 3). В дополнение, была измерена реальная производительность графического процессора для нашего алго-

ритма, результат представлен на рис. 4. На рис. 2-4 представлены результаты расчетов с двойной точностью. Все представленные расчеты также были выполнены с одинарной точностью.

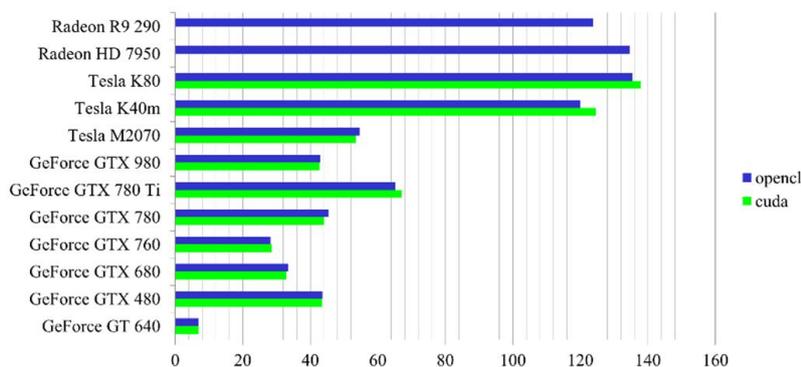


Рис. 4. Производительность в ГФлопсах для рассматриваемого алгоритма

7. Заключение

В данной работе удалось получить ускорение по сравнению с CPU на одном графическом устройстве – в 55 раз на GeForce GTX 780 Ti с одинарной точностью и в 44 раза на Tesla K80 с двойной точностью (рис. 2). На этих устройствах также было достигнуто и наибольшее значение производительности – 460 GFLOPS и 138 GFLOPS (рис. 4).

Была продемонстрирована эффективность использования технологии OpenCL для решения такого рода задач. Также стоит заметить, что данная технология применима на более широком классе устройств, чем CUDA, но при этом производительность отличается несущественно.

Список литературы

- LeVeque R. J.* Finite volume methods for hyperbolic problems. — Cambridge university press, 2002. — Vol. 31.
- Golubev, V. I., Khokhlov, N. I., Petrov, I. B., Zhdanov, M. S., Velikhov, V. E.* Novel Approach to Modelling Elastic Wavefields in Fractured Media // 77th EAGE Conference and Exhibition 2015. — 2015.
- Petrov I. B., Khokhlov N. I.* Modeling 3D seismic problems using high-performance computing systems // Mathematical Models and Computer Simulations. — 2014. — Vol. 6, No. 4. — P. 342–350.
- Castro C.E., Behrens J., Pelties C.* CUDA-C implementation of the ADER-DG method for linear hyperbolic PDEs // Geoscientific Model Development Discussions. — 2013. — Vol. 80, No. 3. — P. 3743–3786.
- Khanna G.* High-Precision Numerical Simulations on a CUDA GPU: Kerr Black Hole Tails // Journal of Scientific Computing. — 2013. — Vol. 56, No. 2. — P. 366–380.
- Muratov M. V., Petrov I. B., Sannikov I. V., Favorskaya A. V.* Grid-characteristic method on unstructured tetrahedral meshes // Comput. Math. and Math. Phys. — 2014. — Vol.54, No.5. — P. 837–847

Applying OpenCL technology for modeling seismic processes with grid-characteristic method

A. M. Ivanov, N. I. Khokhlov

Moscow Institute of Physics and Technology (State University)
9, Institutskiy lane, Dolgoprudny, 141701, Russia

The paper deals with the use of OpenCL technology for modeling seismic processes. We solve the problem of two-dimensional seismic wave propagation. We find numerical solution of elastic wave equation. To find a solution we use grid-characteristic method. It is applicable if the system of equations is hyperbolic, and therefore, the results of this study are applicable to other problems, which are reduced to the solution of hyperbolic systems. We have compared the performance of a numerical method on the central processor (CPU) and graphics processor (GPU). The GPU tests using OpenCL technology compared with the same tests using CUDA technology. The performance of the algorithm using OpenCL technology doesn't differ significantly from the performance of CUDA implementation, but OpenCL allows writing programs for a wider range of graphics devices. Calculations were carried out both with double-precision and single-precision accuracy. Compared with the single-core CPU Intel Xeon E5-2697 we managed to get the acceleration of 44 times when using the graphics card GeForce GTX 780 Ti with single-precision calculations. This performance was equal to 460 Gflops. For calculations with double-precision accuracy we obtained acceleration up to 55 times on the GPU Tesla K80. The resulting performance was 138 GFlops. Greatest double-precision performance was achieved on the Tesla series GPUs, greatest single-precision – on the Geforce GPUs. However, the Radeon GPUs show average results when compared to other cards. Besides performance and acceleration comparisons with CPU, we determined percentage of peak performance theoretically achievable on each graphics device. Using OpenCL technology for the realization of grid-characteristic method allowed utilizing up to 32% of the peak performance of the graphics card, which is much higher than can be achieved for most implementations of computational algorithms.

Keywords: mathematical modeling, grid-characteristic, hyperbolic, parallel programming, GPU, OpenCL

The study was financially supported by RFBR as part of a research project number 15-07-01931 A.

© 2016 A. M. Ivanov, N. I. Khokhlov