

СОВРЕМЕННЫЕ ВОЗМОЖНОСТИ ОРГАНИЗАЦИИ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ С ИСПОЛЬЗОВАНИЕМ ПЕРСОНАЛЬНЫХ ЭВМ

А.Е. Розенбаум, А.В. Глазков

MODERN FEATURES OF THE HIGH PERFORMANCE COMPUTING ORGANIZATION USING A PERSONAL COMPUTERS

A.E. Rosenbaum, A.V. Glazkov

Рассматриваются современные возможности организации высокопроизводительных вычислений на базе ПЭВМ с применением стандартных аппаратных графических ускорителей и соответствующих инструментальных программных средств с использованием сформировавшихся в данной области промышленных стандартов. Приводятся экспериментальные результаты по достигаемой вычислительной производительности на некоторых видах алгебраических задач и задач имитационного моделирования в сравнении со стандартными реализациями на базе современных процессоров общего назначения. Полученные результаты позволяют предположить достижимость прироста вычислительной мощности до 1–2 порядков на определённых видах задач с использованием типовых аппаратных средств.

Ключевые слова: высокопроизводительные вычисления, графический ускоритель.

The article deals with the features of organizing of modern high performance computing based on PCs with standard hardware graphics accelerators and related software tools using mature industrial standards. Experimental results for computing performance on some types of algebraic problems and problems of simulation in comparison with standard implementations based on modern general-purpose processors are achieved. These results suggest the accessibility of computing power increase up to 1–2 orders of magnitude for certain types of problems with the use of standard hardware.

Keywords: high performance computing, graphics accelerator.

В процессе машиностроительного проектирования в ОКБ «Новатор» возникают задачи, решение которых требует проведения значительного объёма вычислений. Например, задачи из таких областей, как термодинамика, механика сплошных сред, газодинамика, гидродинамика, оптимизация управления, имитационное моделирование систем, обработка сигналов, для решения которых используется как стандартное программное обеспечение, так и специализированное программное обеспечение собственной разработки. Ввиду ограниченности доступных вычислительных ресурсов компьютеров общего назначения (персональных ЭВМ) и относительной малодоступности специализированных вычислительных систем (большие ЭВМ,

суперЭВМ), зачастую приходится жертвовать точностью при поиске допустимых решений (сокращать глубину и ширину области поиска решения, уменьшать количество итераций, пользоваться упрощёнными моделями и приближёнными функциями и т. д.).

На сегодняшний день развитие аппаратных возможностей персональных ЭВМ и инструментальных средств разработки сделало возможным разработку и эксплуатацию программ вычислительного характера с использованием массового параллелизма в рамках одной персональной ЭВМ. Основным стимулом для этого процесса послужил практически достигнутый современной индустрией технологический потолок для дальнейшего на-

Розенбаум Александр Евгеньевич – ведущий инженер-программист ОАО «ОКБ «Новатор», г. Екатеринбург; main@okb-novator.ru

Глазков Андрей Валерьевич – инженер-программист ОАО «ОКБ «Новатор», г. Екатеринбург; main@okb-novator.ru

Rosenbaum Aleksandr Evgenievich – leading software engineer of OSC «EDB «Novator», Yekaterinburg; main@okb-novator.ru

Glazkov Andrey Valerievich – software engineer of OSC «EDB «Novator», Yekaterinburg; main@okb-novator.ru

ращивания производительности систем путём увеличения тактовых частот и усложнения внутреннего конвейера процессоров.

В части аппаратных средств можно выделить появление в последние годы высокопроизводительных многоядерных и многопоточных центральных процессоров с перспективой значительного роста числа вычислительных ядер в обозримом будущем. В частности, на сегодняшний день реально доступными стали 4-ядерные 8-поточные центральные процессоры компаний Intel и AMD и появились сообщения о выводе на рынок 6-ядерных, а в ближайшей перспективе 12- и даже 16-ядерных центральных процессоров. Эти же компании-производители прилагают значительные усилия в части создания адекватных инструментальных средств для разработки программного обеспечения с использованием высоких степеней параллелизма (компиляторы, отладчики, системные библиотеки), а также принимают активное участие в процессах выработки и распространения соответствующих стандартов (например, OpenMP). Многоядерные и многопроцессорные решения, являясь процессорными архитектурами общего назначения с унифицированным доступом к памяти, позволяют реализовывать алгоритмы высокой степени сложности с произвольным доступом к любым данным, находящимся в оперативной памяти процесса, обладают развитыми средствами организации межпоточного взаимодействия и защиты данных. Это открывает возможности по реализации на таких системах параллельного решения задач с использованием как параллелизма на уровне данных (значительный объём данных параллельно обрабатывается по одному и тому же алгоритму), так и параллелизма по задачам (одновременное решение нескольких, возможно разных, независимых друг от друга задач с последующим объединением результатов).

Одновременно с развитием современных процессоров в последние годы наблюдался значительный рост возможностей видеоадаптеров, в основном в целях повышения качества развлекательной трёхмерной графики. Этот прогресс привёл к фактическому появлению специализированных вычислительных устройств (графических процессоров), обладающих потенциалом по достижению чрезвычайно высоких степеней параллелизма.

Доступные на сегодняшний день на розничном рынке решения содержат от 800 (ATI HD4870) до 1600 (ATI HD5870) и даже до 3200 (ATI HD5970) вычислительных блоков. Компания NVIDIA также анонсировала выпуск в ближайшее время примерно эквивалентных графических систем на базе архитектуры с кодовым названием Fermi. Потенциальные возможности, скрытые в этих устройствах, не могли не заинтересовать разработчиков приложений, требующих интенсивных вычислений, и уже в течение нескольких лет на рынке появляются примеры программ, задейст-

вующих этот ресурс. В качестве примеров можно привести задачи по обработке двухмерной графики, конвертирования видеопотоков, а также, что особенно интересно, задачи, связанные с моделированием физических систем. Использование ресурсов графических ускорителей для решения таких задач позволяет добиваться ускорения в десятки, а в отдельных случаях и в сотни раз. Развитие этого перспективного направления сдерживалось высокой сложностью достаточно низкоуровневого программирования и существованием несовместимых между собой систем разработки для графических процессоров разных производителей. Компания NVIDIA для этих целей продвигала свою технологию с кодовым названием CUDA, а компания ATI своё конкурирующее и несовместимое решение продвигала под названием ATI Stream. На базе обеих технологий появились достаточно эффективные примеры программ, демонстрирующие чрезвычайно высокий потенциал данной технологии.

К счастью, во второй половине 2009 года намечился значительный прогресс в области унификации технологий использования возможностей графических процессоров для решения вычислительных задач общего назначения.

В частности, организацией Kronos Group, широко известной своей деятельностью по разработке стандарта OpenGL для систем трёхмерной графики, в октябре 2009 г. выпущена окончательная спецификация версии 1.0 широко поддерживаемого индустрией стандарта OpenCL для программирования массовых параллельных вычислений. На сегодняшний день поддержка этого стандарта уже реализована компаниями ATI и NVIDIA, причём на разных платформах (в том числе, кроме Microsoft Windows, и в средах Linux и MacOS X Snow Leopard). Кроме того, компания ATI, являясь подразделением компании-производителя процессоров AMD, в своей реализации OpenCL обеспечила возможность исполнения неизменного исходного кода и на многоядерных центральных процессорах.

Практически одновременно компанией Microsoft была выпущена новая версия операционной системы – Microsoft Windows 7, включившая в себя библиотеку DirectX 11, которая содержит в себе стандартизованную подсистему DirectCompute для организации вычислений общего назначения на совместимых графических процессорах. Для работы программ, использующих технологию DirectCompute, требуется наличие графического процессора с аппаратной поддержкой возможностей DirectX 10.1 и операционных систем Microsoft Windows Vista либо Microsoft Windows 7. Данная технология является более низкоуровневой с точки зрения программирования и на видеопроцессорах предыдущего поколения позволяет добиваться более высокого быстродействия. Область её применимости ограничена только системами на базе Microsoft Windows.

К сожалению, графические процессоры, являясь специализированными вычислителями, не дают возможность эффективно решать любые вычислительные задачи. Слабым местом являются такие факторы, как ограниченная поддержка вычислений с использованием чисел с плавающей запятой двойной точности, невозможность произвольного доступа к памяти большого объёма, необходимость передавать исходные данные и результаты вычислений между основной оперативной памятью ПЭВМ и собственной оперативной памятью графического процессора через относительно медленную внутреннюю шину, что приводит к значительным накладным расходам. Также есть весьма серьёзные ограничения на внутреннюю сложность используемых алгоритмов обработки данных. Однако имеющихся возможностей вполне достаточно для эффективной реализации простых задач линейной алгебры, а соответственно и систем моделирования физических процессов с большим количеством частиц, либо ячеек моделирования (при использовании метода конечных элементов). Существуют весьма эффектные примеры использования графических процессоров в решении криптографических задач.

Проведённые нами эксперименты продемонстрировали реально достижимый результат. Так, например, задача умножения двух квадратных матриц размерностью 2000×2000 элементов с одинарной точностью на довольно дорогостоящей системе на базе 4-ядерного 8-поточного процессора Intel Core i7-950 (3,06 GHz) выполняется с использованием параллельной многопоточной реализации примерно 3,7 с. Возможно, что при дальнейшей оптимизации этот показатель удастся понизить до 2 с. При этом выполняется $16 \cdot 10^9$ степе-

ни операций с плавающей запятой, то есть достигается производительность порядка 8 GFLOPS. Исполнение этой же задачи с использованием технологии OpenCL на графической карте ATI 4870 (800 конвейеров на частоте 750 MHz) производится примерно за 138 мс (вместе с передачей исходных данных в память графического процессора и извлечением результатов), что даёт производительность около 115 GFLOPS.

Применение технологии DirectCompute на этой же графической карте для решения задачи имитации гравитационного взаимодействия системы из 10 000 тел даёт цифры более 700 GFLOPS, причем в это же самое время часть ресурсов тратится ещё и на визуализацию результата. Пример, связанный с имитацией такой системы из 30 000 тел, даёт цифры, превышающие 800 GFLOPS.

Измерения скорости показывают, что при применении графических карт класса ATI HD5970 можно ожидать дополнительного 4–8-кратного роста производительности. Правда, и стоимость таких адаптеров растёт примерно с этой же скоростью, но всё равно остается сопоставимой со стоимостью системы на базе процессора Intel Core i7.

Проведённые эксперименты демонстрируют реальный технологический потенциал для существенного (до 1–2 порядков) повышения возможностей решения ресурсоёмких вычислительных задач на ПЭВМ, правда, это связано со значительно возрастающей трудоёмкостью процесса разработки. Остаётся надеяться на скорое появление соответствующих системных библиотек, эффективно реализующих основные алгоритмы численных научных расчётов с использованием технологий OpenCL и DirectCompute. Первые прототипы таких библиотек уже начинают появляться.

Поступила в редакцию 12 октября 2010 г.