

А.А. Колпаков

*Муромский институт (филиал) Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых
602264, г. Муром, Владимирской обл., ул. Орловская, 23
E-mail: desT.087@gmail.com*

Увеличение производительности гетерогенных многопроцессорных вычислительных систем

Известно, что повышение эффективности вычислительных компьютерных систем осуществляется в зависимости от организации процесса решения задач. В общем случае задачи представляются параллельными программами и описываются рядом параметров, в числе которых: количество ветвей, ранг необходимой подсистемы, время решения и т.п. Режим функционирования высокопроизводительных вычислительных систем формируется мультипрограммным методом или в некоторых вычислительных компьютерных системах используется частичное применение вычислительных модулей, что в недостаточной степени обеспечивает повышение производительности вычислений.

В связи с этим возникает задача разработки методов повышения производительности компьютерных систем на основе модели архитектуры с использованием дополнительных вычислительных производительных модулей или с использованием однородных модулей на графических процессорах. Основной задачей повышения производительности такой вычислительной системы является решение проблемы принятия решений о переносе операций вычислений на специализированные вычислительные модули и кэшировании данных, что требует исследований и разработки соответствующих алгоритмов.

Для рассмотрения особенностей обобщенной архитектуры специализированных вычислительных модулей и их взаимодействия с центральным процессором была разработана и исследована структурная схема архитектуры гетерогенной многопроцессорной вычислительной системы. Базовыми структурными элементами специализированных вычислительных модулей являются спецпамять (SpRAM), в которой отдельно можно выделить память констант и глобальную память, и множество мультипроцессоров. Чтобы обработать данные на специализированных вычислительных модулях, необходимо передать их из оперативной памяти компьютера в SpRAM. Связку «центральный процессор – графический процессор» можно отнести к модели с общей памятью. Основной моделью с общей памятью является модель PRAM (parallel random-access machine) – машина с параллельным произвольным доступом. Она является абстрактной идеализированной моделью параллельной синхронной машины с разделяемой общей памятью, которая использует допущения, приведенные ниже:

- количество процессоров (q) в машине не ограничено;
- каждый процессор имеет равнозначный доступ к любой ячейке общей памяти, размер которой не ограничен;
- отсутствует конкуренция по ресурсам;
- процессоры работают в режиме MIMD, но в частном случае может использоваться режим SIMD.

Все процессоры исполняют инструкции синхронно, причем выполнение любой инструкции занимает ровно 1 такт, называемый шагом PRAM-машины.

Из схемы, приведенной на рис. 1, можно отметить, что PRAM модель может быть применена к многопроцессорной системе с учётом следующих уточнений и дополнений:

- все процессоры могут одновременно считывать данные из разделяемой памяти, но запись должна быть монопольной, т.к. порядок изменения ячейки разделяемой памяти при обращении на запись из нескольких скалярных процессоров не определён (PRAM – CREW (Concurrent Read, Exclusive Write));
- количество скалярных процессоров в графическом мультипроцессоре ограничено сверху (q_{max} процессоров). Для выполнения большего числа потоков используется система горизонтального параллелизма, аналогичная горизонтальной структуре в модели BSP:

Секция 8. Модели и структуры автоматизированных вычислительных и телекоммуникационных систем

генерируется расписание последовательного исполнения потоков, разбитых на пучки по q_{warp} скалярных процессоров;

- размер разделяемой памяти мультипроцессора ограничен – M_s байт;
- все скалярные процессоры работают с одинаковой скоростью по принципу SIMD со скоростью S_{GPU} элементарных операций в секунду;
- должна иметь место дополнительная операция – обращение к оперативной памяти SpRAM специализированного вычислительного модуля на чтение или запись. Задержка при обращении K определяется количеством элементарных операций, требуемых при обращении к одному числу одинарной точности в глобальной памяти специализированного вычислительного модуля.

Таким образом, на основе модифицированной PRAM-модели, разработан алгоритм повышения производительности параллельных вычислений на специализированных вычислительных модулях, который включает в себя алгоритм принятия решения о переносе вычислений на графический процессор.

Методом оценивания производительности были осуществлены сравнительные экспериментальные исследования разработанного алгоритма. Результаты оценивания алгоритма показывают повышение производительности не менее, чем в 2-4 раза в зависимости от числа исследуемых потоков.

Литература

1. Langdon W.B. A Many Threaded CUDA Interpreter For Genetic Programming // EuroGP. 2010, – С. 146-158.
2. Harding S.L., Banzhaf W. Fast genetic programming on GPUs // LNCS: Proceedings of the 10th European Conference on Genetic Programming. 2007, – Vol. 4445. – С. 99-101.
3. Harding S.L., Banzhaf W. Distributed genetic programming on GPUs using CUDA // WPABA: Proceedings of the Second International Workshop on Parallel Architectures and Bioinspired Algorithms. 2009, – С. 1-10.
4. Колпаков А.А. Оптимизация генетических алгоритмов при использовании вычислений на графических процессорах на примере задачи нулевых битовых векторов // Информационные системы и технологии. – 2013, - №2(76) – С. 22-28
5. Догадина Е.П., Кропотов Ю.А., Суворова Г.П. Математическая модель определение вероятностей системы обслуживания // Радиотехника, 2009. – №11. – С.103-105.
6. Колпаков А.А. Теоретическая оценка увеличения производительности вычислений при распараллеливании процессоров вычислительных систем // В мире научных открытий, 2012. – №1. – С. 51-52.
7. Догадина Е.П., Суворова Г.П., Кропотов Ю. А. Оценка параметров вычислительных процессов при циклическом планировании // Информационные системы и технологии, 2010. – №3(59). – С. 12-19.
8. Колпаков А.А., Кропотов Ю.А. Аспекты оценки увеличения производительности вычислений при распараллеливании процессоров вычислительных систем // Методы и устройства передачи и обработки информации, 2011. – №1(13). – С.124-127.