

Анализ применения технологии Nvidia CUDA при моделировании динамики плоского электронного потока в скрещенных статических электрическом и магнитном полях

Е.А. Шамов, Д.А. Стуров, О.В. Шаповалов

Настоящая работа представляет собой анализ применения технологии Nvidia CUDA при моделировании трехмерной модели движения плоского электронного потока в скрещенных статических электрическом и магнитном полях путем сравнения исследуемой технологии с технологиями распараллеливания вычислений OpenMP, MPI, комбинированной OpenMP+MPI. Эффективность полученных результатов подтверждена содержанием и выводами работы.

В рамках анализа были сделаны следующие выводы:

1. Технология CUDA позволяет приложениям, активно работающим с данными, получать доступ к мощностям обеспечения вычислений на графических процессорах Nvidia.
2. При решении поставленной задачи в данной работе были применены четыре технологии параллельного вычисления: OpenMP, MPI, совместно OpenMP и MPI, Nvidia CUDA.
3. Сравнение результатов применения описанных технологий приводит к решению об оптимальности применения технологии CUDA так как поскольку сравнение технологии CUDA проходило с технологией совместного применения OpenMP и MPI выявлено следующее:

Программа моделирования электронного потока сама по себе является практически полностью параллельной. При повышении числа электронов в потоке эффективность распараллеливания возрастает, а при увеличении числа ядер (при том же числе электронов) эффективность их использования падает. Поэтому с увеличением числа ядер целесообразно рассчитывать большие электронные потоки. Так как задача по моделированию электронных потоков заключается в приближении моделируемого потока к реальному, а соответственно в увеличении числа частиц, то, чтобы повысить точность расчетов и приблизить моделируемый поток к реальному, необходимо увеличить число электронов (размерность задачи). Однако также необходимо сокращать и время расчетов для заданной размерности задачи, что весьма сложно сделать, так как при увеличении числа ядер растет время пересылок информации, а относительное время самих расчетов уменьшается. Поэтому в некоторый момент дальнейшее увеличение числа ядер (при постоянной размерности задачи) не даст существенного уменьшения времени расчета.

Полученные результаты говорят о том, что использование видеокарты для расчета электронных потоков весьма эффективно. Скорость расчета данной задачи на Nvidia GeForce GTX 260 сравнима со скоростью расчета на кластере с пиковой производительностью 0,76 TFlops. С учетом значительного различия в ценовой характеристике, а также сложности организации и настройки кластера использование технологии Nvidia CUDA можно считать на данный момент наиболее оправданным.

Вычислительные эксперименты были проведены на учебном кластере ВолгГТУ на кафедры «ЭВМ и систем». Кластер построен на базе одно- и двухпроцессорных узлов с четырехядерными процессорами Intel Xeon/Quad и операционной системы Microsoft HPC2008, максимальное количество процессорных ядер - 72.

Видеокарта начинает существенно проигрывать кластеру на больших задачах длительностью несколько часов, на малых же размерностях Nvidia CUDA даже превосходит кластер. Это объясняется большой долей пересылок при расчете малой задачи на кластере, в то время как с помощью CUDA легче и быстрее достигается максимальная эффективность использования видеокарты. В качестве дальнейших перспектив развития программы видится одновременное применение для расчетов нескольких видеокарт, расположенных как на одной рабочей станции, так и на нескольких с использованием технологии MPI. В результате скорость решения данной задачи моделирования электронного потока должна возрасти примерно в $2N/3$ раз, где N - количество видеокарт в системе. Ускорение будет непропорционально N из-за накладных расходов, связанных с необходимостью разбиения задачи и пересылки данных.