

Реферат

магистерской аттестационной работы

на тему:

“Исследование алгоритмов работы с разреженными матрицами в среде CUDA ”

Горбика Александра Викторовича

Актуальность работы

Скорость работы алгоритмов с разреженными матрицами является критичной для многих приложений, в частности для схемотехнических пакетов, работающих с матрицами больших размеров и малой величиной заполнения. Операция умножения разреженной матрицы на вектор-столбец лежит в основе многих алгоритмов, характерных для моделирования и анализа работы электронных схем. В работе с разреженными матрицами явно выраженный параллелизм по данным - обработка массива элементов может происходить независимо. Для эффективного использования заложенного параллелизма целесообразно использовать массивно-параллельные архитектуры графических карт.

Технология CUDA предоставляет мощный и удобный интерфейс для вычислений общего назначения на графических картах. С целью ускорения работы алгоритмов необходимо учитывать особенности форматов представления разреженных матриц и самой архитектуры. На практике использование CUDA дает ускорение отдельных алгоритмов в 10-30 раз. Поэтому разработка и исследование эффективных методов использования вычислительных мощностей, в частности, CUDA является весьма актуальной проблемой.

Цель работы

Целью работы является разработка алгоритмов работы с разреженными матрицами с использованием технологии вычислений общего назначения на графических

картах CUDA и исследование возможности использования ускоренных реализаций алгоритмов для приложений схемо технического моделирования.

Задачи, решаемые в работе

1. Исследование особенностей работы с массивно-параллельной архитектурой CUDA.
2. Исследование возможности ускорения основных алгоритмов работы с разреженными матрицами с использованием технологии CUDA.
3. Реализация алгоритма умножения разреженной матрицы на вектор на стороне графического ускорителя для форматов DIA, ELL, COO, CSR и ALLTED.
4. Реализация метода решения разреженной системы линейных уравнений на стороне графического ускорителя.
5. Исследование возможности использования реализаций алгоритмов для приложений.

Достигнутые результаты

Решив задачи, поставленные в работе, автор защищает:

1. результаты исследования особенностей работы с массивно-параллельной архитектурой CUDA;
2. результаты исследования возможности ускорения основных алгоритмов работы с разреженными матрицами с использованием технологии CUDA;
3. реализации алгоритмов умножения разреженной матрицы на вектор на стороне графического ускорителя для форматов DIA, ELL, COO, CSR и ALLTED;

4. реализацию итерационного метода решения разреженной системы линейных уравнений на стороне графического ускорителя;
5. результаты исследования возможности использования реализаций алгоритмов для приложений.

Научная новизна работы

Научная новизна работы заключается в том, что:

1. исследованы основные особенности работы с технологией CUDA и выявлены возможности ее использования в схемотехнических пакетах
2. исследовано влияние выбора формата представления разреженных матриц на скорость работы алгоритмов с ними. Разработан программный модуль для тестирования работы основных алгоритмов работы с разреженными матрицами в среде CUDA;
3. математическая модель взаимодействия GPU и CPU;
4. разработаны функции ядра для работы на графической карте, которая занимается умножением матрицы на вектор формата, который использует ППП ALLTED.

Практическая ценность работы

Практическая ценность работы заключается в том, что:

1. экспериментально исследована и доказана эффективность использования среды CUDA для прикладных пакетов схемотехнического моделирования;
2. экспериментально исследована реализация умножения матрицы на вектор формата использует ППП ALLTED.

Выводы

1. Исследованы особенности работы с массивно-параллельной архитектурой CUDA. Рассмотрены основные проблемы, связанные с эффективным использованием графического оборудования.
2. Исследованы возможности ускорения основных алгоритмов работы с разреженными матрицами с использованием технологии CUDA.
3. Реализованы алгоритмы умножения разреженной матрицы на вектор на стороне графического ускорителя для форматов DIA, ELL, COO, CSR и ALLTED. Исследованы проблемы использования указанных форматов на стороне графической карты.
4. Реализован итерационный метод решения разреженной системы линейных уравнений на стороне графического ускорителя. Проведен анализ его исполнения и оптимизированы затраты на пересылку данных.
5. Исследована возможность использования реализаций алгоритмов для приложений, в частности, в пакете схемотехнического моделирования ALLTED с использованием экспортируемой модели Ansys.

Работа содержит 160 с., 23 рисунки, 7 таблиц, 29 источников.

Ключевые слова: МАССИВНО-ПАРАЛЛЕЛЬНАЯ АРХИТЕКТУРА, РАЗРЕЖЕННАЯ МАТРИЦА, CUDA, ВЫЧИСЛЕНИЕ ОБЩЕГО НАЗНАЧЕНИЯ НА ГРАФИЧЕСКИХ КАРТАХ.