

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра дискретной математики и информационных технологий

**МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ПЕРЕНОСА НА СОПРОЦЕССОРАХ
С АРХИТЕКТУРОЙ CUDA**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 271 группы
направления 09.04.01 — Информатика и вычислительная техника
факультета КНиИТ
Кузнецова Максима Олеговича

Научный руководитель
доцент, к. ф.-м. н.

А. Д. Панферов

Заведующий кафедрой
доцент, к. ф.-м. н.

Л. Б. Тяпаев

Саратов 2020

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Кинетическая модель поведения графена во внешнем электрическом поле	5
2 Понятие ОДУ. Задача Коши	6
3 Технология CUDA.....	7
4 Параллельная реализация решения системы ОДУ в двумерном про- странстве параметров с использованием CUDA.....	8
5 Демонстрация разработанного решения в составе программного ком- плекса моделирования поведения графена	9
ЗАКЛЮЧЕНИЕ	13
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	14

ВВЕДЕНИЕ

В настоящее время наращивание аппаратного параллелизма является основным инструментом достижения максимально высокой производительности вычислительных систем. В связи с исчерпанием технологических возможностей повышения тактовых частот универсальных процессоров (CPU), их развитие пошло по пути увеличения количества ядер. При этом увеличение числа аппаратных ядер для CPU со сложной архитектурой относительно дорого и тоже приближается к естественному пределу, определяемому достижимыми параметрами литографии.

В практике научного и промышленного моделирования существует достаточно много вычислительно сложных задач, потенциал параллелизма которых превосходит возможности массово параллельных систем, построенных на CPU. При этом архитектурные возможности ядер таких процессоров избыточны для программ с доминированием однопоточных математических операций. По этой причине оказалось эффективным использование специализированных математических ускорителей (сопроцессоров), в которых приоритет отдаётся количеству доступных для использования ядер с упрощённой архитектурой. Вычислительные системы такого типа, в которых в дополнение к CPU устанавливаются специализированные аппаратные ускорители, принято называть гибридными. При проектировании современных гибридных вычислительных систем наиболее часто предпочтение отдаётся ускорителям компании NVIDIA. Так, в списке 500 наиболее производительных компьютеров (Тор-500) по состоянию на июнь 2019 года суммарная производительность систем с ускорителями от NVIDIA различных моделей достигла 35% суммарной производительности всех вычислительных систем, вошедших в этот список.

Эффективное использование ускорителей требует, однако, специальной адаптации программ в силу их архитектурных отличий от универсальных процессоров. Такая адаптация была основной целью представляемой работы. Она выполнялась для наиболее ресурсоемкого блока системы моделирования процессов переноса заряда в графене, ответственного за решение квантового кинетического уравнения в форме системы обыкновенных дифференциальных уравнений (ОДУ) для двумерного массива параметров, определяющих характер поведения переменных коэффициентов этой системы ОДУ.

Полученное в работе решение было интегрировано в программный ком-

плекс, обеспечивающий детальное воспроизведение функции распределения носителей заряда, формируемой в результате воздействия на моделируемый материал внешним электрическим полем, и предоставляющей полную информацию о результирующем потоке заряда (плотности тока).

Разработанное программное решение тестировалось на модельных задачах для определения и демонстрации его скоростных характеристик. После интеграции его в программный комплекс было выполнена полная процедура моделирования результата действия короткого инфракрасного импульса с линейной поляризацией.

Магистерская работа состоит из введения, 5 разделов, заключения, списка использованных источников и 2 приложений. Общий объем работы – 70 страниц, из них 40 страниц – основное содержание, включая 17 рисунков, список использованных источников информации – 30 наименований.

1 Кинетическая модель поведения графена во внешнем электрическом поле

В первой части работы кратко изложены характеристики модели, приведен явный вид решаемой системы ОДУ, определена зависимость переменных коэффициентов этой системы от параметров моделируемого процесса и времени.

2 Понятие ОДУ. Задача Коши

Во второй части представлены базовые данные по методам численного решения задачи Коши для ОДУ и систем ОДУ. Необходимость включения этого материала обусловлена тем обстоятельством, что в ходе разработки версии программы для сопроцессоров NVIDIA возникли непреодолимые трудности в прямом переносе решателей из библиотеки GSL, использовавшихся на CPU. Пришлось искать альтернативу этой библиотеке и реализовывать процедуры решения с использованием других численных методов. Детали этих проблем и найденные решения описаны далее в ВКР.

3 Технология CUDA

Третья часть работы посвящена описанию программной технологии CUDA, являющейся разработкой NVIDIA и предоставляющей удобные инструменты для написания приложений для ускорителей этой компании.

4 Параллельная реализация решения системы ОДУ в двумерном пространстве параметров с использованием CUDA

В четвертой части излагаются детали разработанного программного решения с описанием использовавшихся инструментов и библиотек.

5 Демонстрация разработанного решения в составе программного комплекса моделирования поведения графена

Пятая часть представляет результаты демонстрации разработанного решения в составе моделирующего программного комплекса.

Демонстрационное моделирование выполнялось с набором параметров, представленных на рисунке [1](#). Этот набор параметров соответствует реалистичному сценарию действия на образец графена относительно длинного (25 периодов волны) инфракрасного линейно поляризованного импульса с частотой 100 ТГц. Плотность энергии в импульсе обеспечивает генерацию поля с максимальной напряженностью в 300000 В/см. Определенные в представленном варианте файла task_q.txt параметры «ширина рассматриваемой области» и «масштаб сплошного покрытия области» обеспечивают построение уже на первом этапе процедуры квадродерева с полным заполнением восьми поколений узлов (включая корневой). За вычетом корневого узла вычисление функции распределения и вспомогательных функций необходимо выполнить для 21844 узлов покрывающей сетки.

0.0027573	Ea1_Амплитуда_поля_1
0.000	Ea2_Амплитуда_поля_2
0.0246	nu1_Частота_1
0.000	nu2_Частота_2
0.0	fi1_Фаза_1
0.0	fi2_Фаза_2
0.00	t01_Время_максимума_1
0.00	t02_Время_максимума_2
25.0	sIgm1_Сигма_1
3.0	sIgm2_Сигма_2
-1200.0	Время_начало
1200.0	Время_конец
0.0	Время_начала_записи_параметров
1200.0	Время_завершения_записи_параметров
1	Число_шагов_по_времени_в_области_записи
0.0	Первая_координата_центра_сетки
0.0	Вторая_координата_центра_сетки
1.0	Ширина_рассматриваемой_области
0.01	Масштаб_сплошного_покрытия_области
0.0001	Минимальный_масштаб_сетки
0.000000000001	Минимальное_значение_для_функции_распределения
0.2	Погрешность_интерполяции

Рисунок 1 – набор реалистических параметров

Результаты вычислительной процедуры представлены на рисунке [2](#). Это как раз область импульсного пространства в диапазоне от -0.5 до 0.5 (что соответствует значению 1.0 для ширины рассматриваемой области). Максимальные отображаемые значения f_1 достигают почти 0.7, но в подавляющем большинстве узлов сетки значения на многие порядки меньше и в используемом линейном варианте цветовой шкалы все отображаются одинаковым синим цветом. Значения из диапазона примерно от 0.05 до 0.7 сосредоточены на окружности в центре рисунка.

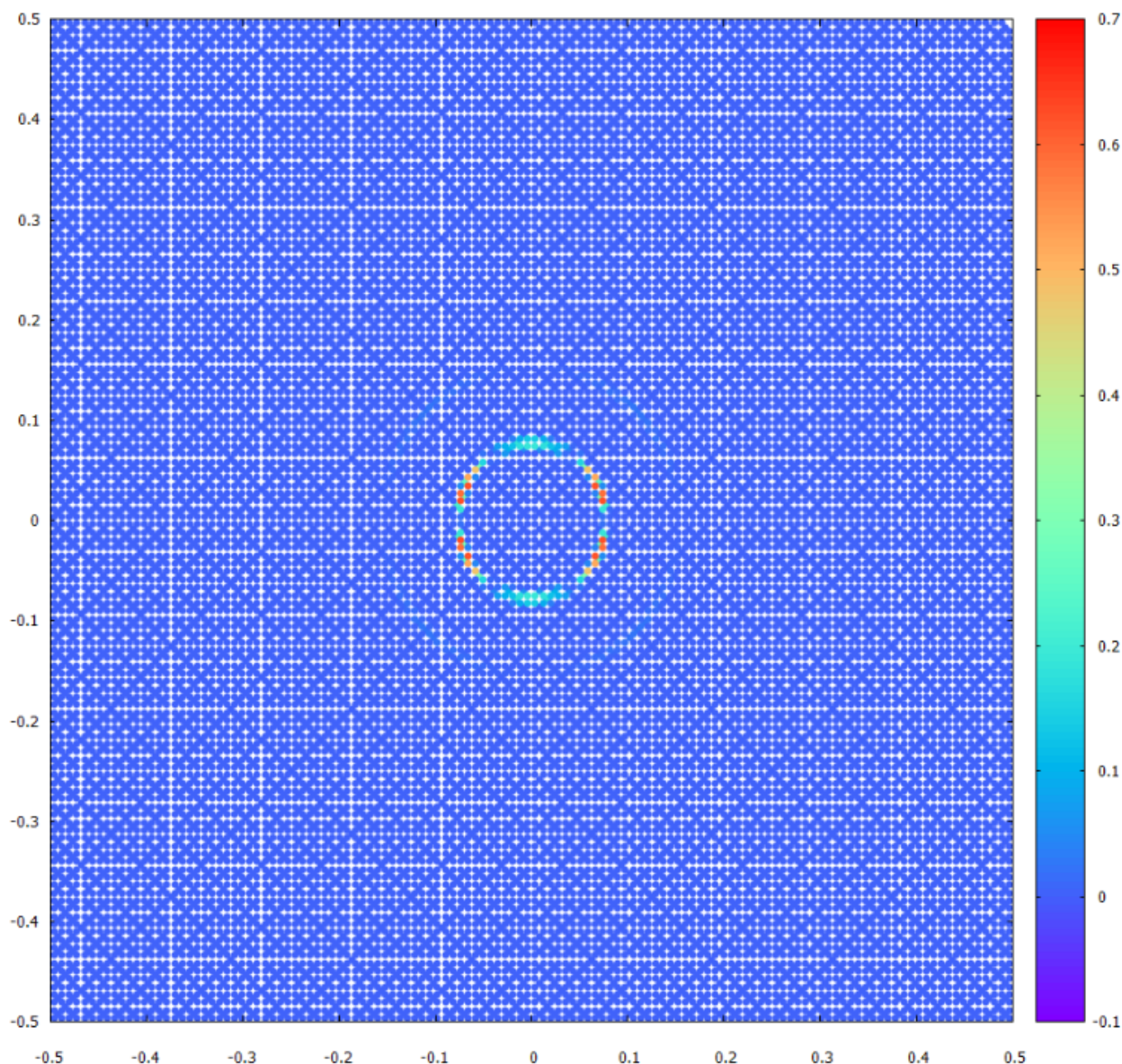


Рисунок 2 – график распределения значений на 1 этапе

После вычисления значений для стартового квадродерева были выполнены три итерации по его достраиванию. На каждой итерации достраивались узлы одного следующего поколения. Но теперь это делалось не путем сплошного квадрирования узлов предыдущего поколения, а только избирательно с анализом локального поведения функции распределения (данная процедура в представляемой работе не рассматривается). На этих итерациях было добавлено 11856, 33700 и 78870 узлов соответственно. Процедура каждой итерации включала вычисление функции распределения и вспомогательных функций для всех добавленных узлов сетки. Это позволило набрать большую статистику по времени счета. Для оценки быстродействия разработанного программ-

ного модуля расчеты дублировались на имеющейся MPI реализации, запущенной на CPU (один процессор, восемь ядер). Результаты представлены на рисунке 3. Они подтверждают эффективность использования ускорителей на задачах такого характера.

Номер этапа (кол-во узлов)	Время выполнения MPI, с	Время выполнения CUDA, с
1 этап (21844 узла)	113.546130	47.7712
2 этап (11856 узлов)	57.067583	49.8807
3 этап (33700 узлов)	181.90521	70.5443
4 этап (78870 узлов)	370.78	122.57

Рисунок 3 – сравнение производительности

Представленные результаты показывают эффективную и корректную работу разработанного программного модуля на масштабной и ресурсоемкой задаче.

ЗАКЛЮЧЕНИЕ

Поставленные в Задании на выполнение ВКР задачи были решены и цели достигнуты. В ходе выполнения работы был изучен большой объём теоретического материала. Он включает обзорную информацию о предметной области, в интересах которой выполнялась работа, основные сведения о математических методах и алгоритмах, использовавшихся при решении поставленной задачи. Но основным объектом теоретической части исследования был инструментальный для разработки программ, способных функционировать на вычислительных ускорителях компании NVIDIA. Это, прежде всего, технология CUDA и вспомогательные инструменты и библиотеки, обеспечивающие эффективное решение задач с её использованием.

Практическая часть работы также была полностью реализована и обеспечено функционирование разработанного решения в составе целевого программного комплекса. Полученные результаты подтвердили правильность принятых программных решений и их реализации.

Выполненная работа сделала возможной эффективное использование гетерогенных вычислительных узлов кластера СГУ для решения исследовательских задач в перспективной области моделирования характеристик новых материалов. И выполнение таких исследований на вычислительных ресурсах большей мощности, использующих ускорители NVIDIA.

Использованные технологии и инструменты соответствуют самым современным тенденциям в развитии глобальной суперкомпьютерной индустрии.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 S.A.Smolyansky, D.V.Churochkin, V.V.Dmitriev, A.D.Panferov and B.Kampfer. "Residual currents generated from vacuum by an electric field pulse in 2+1 dimensional QED models EPJ Web Conf., 138 (2017), 06004.
- 2 Smolyansky, S.A.; Panferov, A.D.; Blaschke, D.B.; Gevorgyan, N.T. Nonperturbative kinetic description of electron-hole excitations in graphene in a time dependent electric field of arbitrary polarization. Particles 2019, 2, 208-230.
- 3 Графен: Самый прочный материал в мире | Журнал Популярная Механика [Электронный ресурс]. URL: <https://www.popmech.ru/science/7873-grafen-samyu-prochnyy-material-v-mire/> (дата обращения 27.05.2020) Загл. с экрана Яз.рус.
- 4 А. В. Маханьков, М. О. Кузнецов, А. Д. Панферов. "Эффективность использования сопроцессоров NVIDIA для моделирования поведения носителей заряда в графене". Подготовлено в журнал Программные системы: теория и приложения. 14 ноября 2019 г.
- 5 1. P.R. Wallace, Phys. Rev. 71, p. 622-634 (1947).
- 6 November 2019 | TOP500 Supercomputer Sites [Электронный ресурс]. URL: <https://www.top500.org/lists/2019/11/> (дата обращения 01.06.2020) Загл. с экрана Яз.англ.