МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**
Кафедра дискретной математики и информационных технологий

# ОБРАБОТКА И ВИЗУАЛИЗАЦИЯ РЕЗУЛЬТАТОВ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ ОТКЛИКА ДВУМЕРНОЙ СИСТЕМЫ НА ВОЗМУЩАЮЩЕЕ ВОЗДЕЙСТВИЕ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 421 группы
направления 09.03.01 — Информатика и вычислительная техника
факультета КНиИТ
Лесняка Дениса Дмитриевича

Научный руководитель
доцент, к. ф.-м. н. _____ А. Д. Панфёров

Заведующий кафедрой
доцент, к. ф.-м. н. _____ Л. Б. Тяпаев

Саратов 2023

# ВВЕДЕНИЕ

Характерной особенностью современного этапа развития человеческой цивилизации является экспоненциальный рост объёма генерируемой и обрабатываемой информации. Это является прямым следствием успехов в области построения мощных вычислительных систем, развития телекоммуникаций, информационных технологий в целом. Объёмы информационных потоков давно превзошли возможности человеческого мозга по их обработке и без соответствующих аппаратных и программных инструментов для их анализа были бы просто бесполезны.

Обработка и анализ больших объёмов данных к настоящему времени стали обязательными инструментами бизнеса, политики и других направлений деятельности. Знание соответствующих инструментов, методов, их возможностей является обязательным для деятельности и успеха в современном мире. Тем не менее, осознание проблемы и формальное введение термина «Big Data» (Большие данные) произошло в науке, когда быстро растущие возможности информационных технологий были использованы в тех областях, где достижение новых результатов требовало обработки чрезвычайно большого количества исходных данных (геномика, физика высоких энергий, астрономия и т.п.).

Развитие численных методов моделирования поведения достаточно сложных физических систем также породило проблему генерации больших объёмов промежуточных данных и необходимость применения к ним современных технологий обработки и анализа. Целью представляемой работы является разработка инструментов для анализа результатов численного моделирования отклика двумерной системы, воспроизводящей свойства графена, на возмущающее воздействие в форме внешнего электрического поля высокой интенсивности. Рабочая станция высокой производительности с использованием математической модели процесса в течении часа генерирует порядка ста миллионов значений, требующих дальнейших обработки. Полная процедура моделирования занимает несколько часов. Полученные массивы промежуточных данных необходимо проанализировать и представить в приемлемой визуальной форме.

Для достижения поставленной цели было необходимо решить ряд задач. Так, были изучены различные инструменты. В частности, была проведена подробная работа с библиотекой NumPy, которая предоставляет эффективные структуры данных и функции для работы с многомерными массивами. Для

визуализации данных была использована библиотека Matplotlib, которая предоставляет широкие возможности для создания различных типов графиков и диаграмм. Визуализация данных с помощью Matplotlib позволяет наглядно представить результаты анализа и облегчает восприятие информации. Для построения 3D графиков был использован модуль Pyplot из библиотеки Matplotlib. Этот модуль позволяет создавать 3D графики, которые могут быть полезны при анализе данных с несколькими параметрами или при визуализации результатов моделирования сложных систем. С помощью Pyplot можно создавать поверхности, контуры и другие 3D графики, настраивать оси и внешний вид графиков, добавлять цветовые карты и многое другое. Для загрузки и обработки данных была использована библиотека Pandas. С помощью Pandas можно считывать данные из различных источников, выполнять операции по фильтрации, сортировке и группировке данных, а также проводить агрегирование и статистический анализ. Библиотека Pandas значительно упрощает процесс подготовки и обработки данных перед анализом, что позволяет сэкономить время и упростить работу с большими объёмами информации.

В процессе изучения и применения указанных инструментов и методов, были освоены процедуры и алгоритмы для эффективного анализа результатов численного моделирования. Они включают в себя обработку массивов данных, вычисление статистических характеристик, визуализацию результатов в различных форматах и возможность создания анимации моделируемого процесса.

В данной работе содержится 4 главы, а именно ”обработка больших массивов данных”, ”инструменты”, ”визуализация и анимация результатов моделирования отклика на возмущающее воздействие” и ”обработка данных и их представление в форме временных рядов”.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

**В первой главе** рассматривается понятие больших данных и их важность в современном мире. Большие данные представляют собой обширные и сложные массивы информации, слишком большие для традиционных систем обработки данных. Они включают структурированные, полуструктурированные и неструктурированные данные из разных источников, таких как социальные сети и мобильные устройства. Рост объема и сложности данных привел к созданию программной среды Hadoop в 2005 году, что отметило начало эры больших данных. Организации используют большие данные для анализа клиентского поведения, оптимизации бизнес-процессов и принятия обоснованных решений. Объем, скорость и разнообразие данных являются ключевыми характеристиками больших данных. Обработка больших данных может быть пакетной или в реальном времени, а также включает добычу данных и аналитику данных [1].

Различают несколько типов обработки больших данных. Пакетная обработка осуществляется путем обработки данных пакетами или группами в определенные промежутки времени. Этот тип обработки подходит для автономного анализа без необходимости немедленных результатов. Обработка в реальном времени, напротив, позволяет обрабатывать данные по мере их получения и получать результаты немедленно. Добыча данных включает обнаружение закономерностей и идей в больших наборах информации с использованием статистических алгоритмов и алгоритмов машинного обучения. Аналитика данных включает анализ данных для извлечения выводов и принятия обоснованных решений. Для обработки больших данных применяются специализированные инструменты, такие как Hadoop, Apache Spark и базы данных NoSQL [2]. Они позволяют параллельно обрабатывать и анализировать данные, а также масштабировать вычислительные ресурсы. Python является популярным языком программирования для обработки и анализа данных, и его библиотеки, такие как NumPy, Pandas и Matplotlib, предоставляют мощные инструменты для работы с данными и визуализации. Большие данные имеют широкий спектр применений в различных отраслях, включая бизнес, здравоохранение, государственный сектор и научные исследования. Аналитика больших данных может помочь в улучшении взаимодействия с клиентами, мониторинге здоровья, прогнозировании стихийных бедствий и научных открытиях [3].

**Во второй главе** рассматриваются ключевые инструменты и библиотеки,

используемые для обработки и анализа больших данных. Они предоставляют мощные возможности для работы с массивами данных, визуализации и выполнения аналитических операций. Некоторые из наиболее популярных инструментов в этой области включают NumPy, Matplotlib, Pandas и Pyplot.

NumPy (Numerical Python) – это фундаментальная библиотека для научных вычислений на языке Python. Она предоставляет эффективные структуры данных, такие как многомерные массивы и матрицы, а также функции для выполнения различных математических операций. NumPy позволяет обрабатывать большие объемы данных и проводить вычисления с высокой скоростью, что делает его незаменимым инструментом для работы с большими данными [4].

Matplotlib – это библиотека визуализации данных, которая предоставляет разнообразные инструменты для создания графиков, диаграмм и других визуализаций. Она позволяет представлять данные в наглядной форме, что облегчает их анализ и визуальное представление результатов. Matplotlib широко используется в области аналитики данных, научных исследований и визуализации информации [5].

Pandas – это мощная библиотека для обработки и анализа данных в Python. Она предоставляет высокоуровневые структуры данных, такие как DataFrame и Series, которые позволяют легко работать с табличными данными. Pandas предоставляет функции для чтения и записи данных из различных источников, манипулирования и очистки данных, а также выполнения агрегирования и аналитических операций [6].

Pyplot – это модуль библиотеки Matplotlib, который предоставляет простой интерфейс для создания графиков и диаграмм в Python. Он облегчает процесс создания различных типов графиков, включая линейные графики, столбчатые диаграммы, круговые диаграммы и многое другое. PyPlot предоставляет гибкие настройки для визуализации данных и позволяет создавать высококачественные графические представления [7].

Эти инструменты и библиотеки являются незаменимыми для работы с большими данными в Python. Они обеспечивают эффективные способы обработки, визуализации и анализа данных.

**В третьей главе** были представлены результаты применения рассмотренных выше инструментов к массиву данных, полученному при моделировании отклика двумерной системы на возмущающее воздействие. Такой двумерной

системой является лист графена, а в качестве внешнего возмущения выступает электрическое поле лазерного импульса инфракрасного диапазона. Сама процедура моделирования не является предметом рассмотрения. Результатом моделирования является набор из трех функций трех переменных: функция распределения $f(p_1, p_2, t)$ и вспомогательные функции $u(p_1, p_2, t)$, $v(p_1, p_2, t)$ (в программном коде они обозначены как $f_1$, $f_2$, $f_3$ соответственно). Они определены на конечном множестве точек в пространстве $(p_1, p_2, t)$ и на первом этапе необходимо было составить общее представление об их поведении в этом пространстве. Основной задачей данной главы стояла визуализация и анимация функции распределения $f(p_1, p_2, t)$ из результатов моделирования. Выполнение этой задачи осложнялось тем, что предоставленные данные имели очень большой объем, а также невозможно было визуально представить значения функции в трехмерной области. В рассматриваемом случае естественным способом понижения размерности является фиксация значений времени $t$ и рассмотрение области двумерного пространства $(p_1, p_2)$ в каждый момент времени отдельно. Это позволило воспользоваться инструментами для 3D визуализации. Так как значений параметра момента времени в рассматриваемом файле 502, то было принято решение объединить графики функции $f(p_1, p_2, t)$ и создать анимацию. На рисунке 1 показан промежуточный кадр из анимации 3D графика функции $f(p_1, p_2, t)$.

Далее необходимо было рассмотреть область максимального значения функции $f(p_1, p_2, t)$. Был выбран максимальный момент времени и также отрисован 3D график. После этого требовалось построить график типа density plot или же график плотности. График плотности с использованием цветовой шкалы позволяет визуализировать трехмерные данные на двумерной плоскости. Такие графики особенно полезны для визуализации функций двух переменных, где одна переменная отображается по горизонтальной оси, другая переменная – по вертикальной оси, а значения функции отображаются цветом. Например, график плотности может показывать распределение температуры на географической карте, где цвета представляют различные уровни температуры в разных регионах [8]. Кроме того, необходимо было также построить анимацию для данного типа графика. На рисунке 2 показан промежуточный кадр анимации графиков типа density plot.

**В четвертой главе** рассматривается обработка данных и их представле-

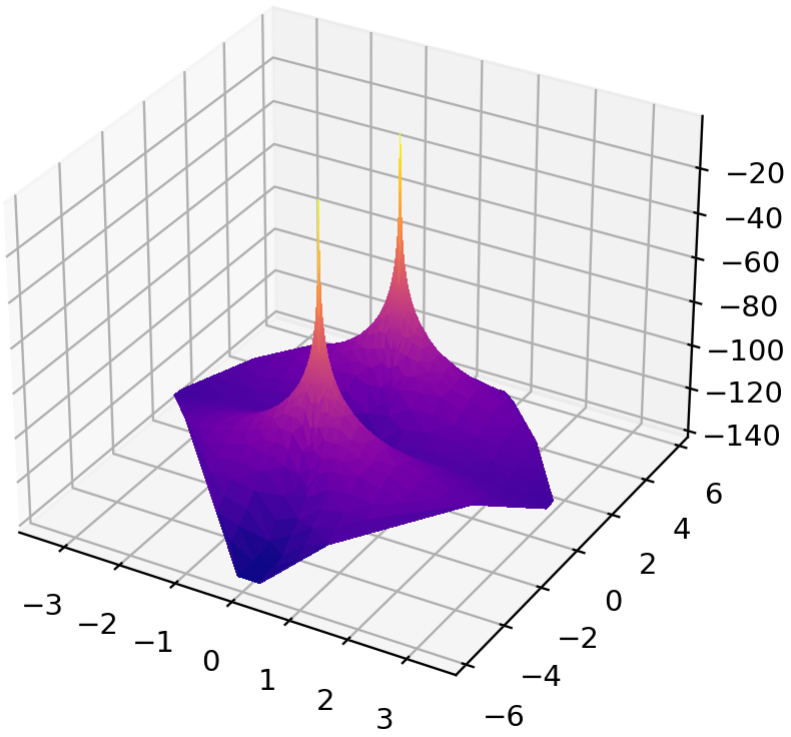


Рисунок 1 – Промежуточный кадр анимации 3D графиков для $10lg * f(p_1, p_2, t)$

ние в форме временных рядов. Конечной целью моделирования любого физического процесса является воспроизведение его характеристик, доступных для измерения в реальных условиях. Рассмотренная выше функция распределения $f(p_1, p_2, t)$ и вспомогательные функции $u(p_1, p_2, t)$, $v(p_1, p_2, t)$ содержат всю необходимую информацию о поведении моделируемой системы, но сами по себе не доступны измерению в экспериментах. Поэтому решаемой задачей являлось получение на основе «сырых» результатов моделирования значений наблюдаемых параметров. Первый из них – поверхностная плотность носителей заряда, то есть количество электронов в возбужденных состояниях на единичной площади образца. Для заданного момента времени эта величина определяется интегралом вида

$$\rho(t_i) = 8 \int \frac{d^2p}{(2\pi)^2} f_1(p_1, p_2, t_i), \tag{1}$$

где интегрирование выполняется по области определения $f(p_1, p_2, t)$ в импульсном пространстве. Еще два наблюдаемых параметра – поверхностные плотности токов свободных зарядов (ток проводимости) и поляризационного тока. Поскольку поверхностные токи являются двумерными векторами, всего

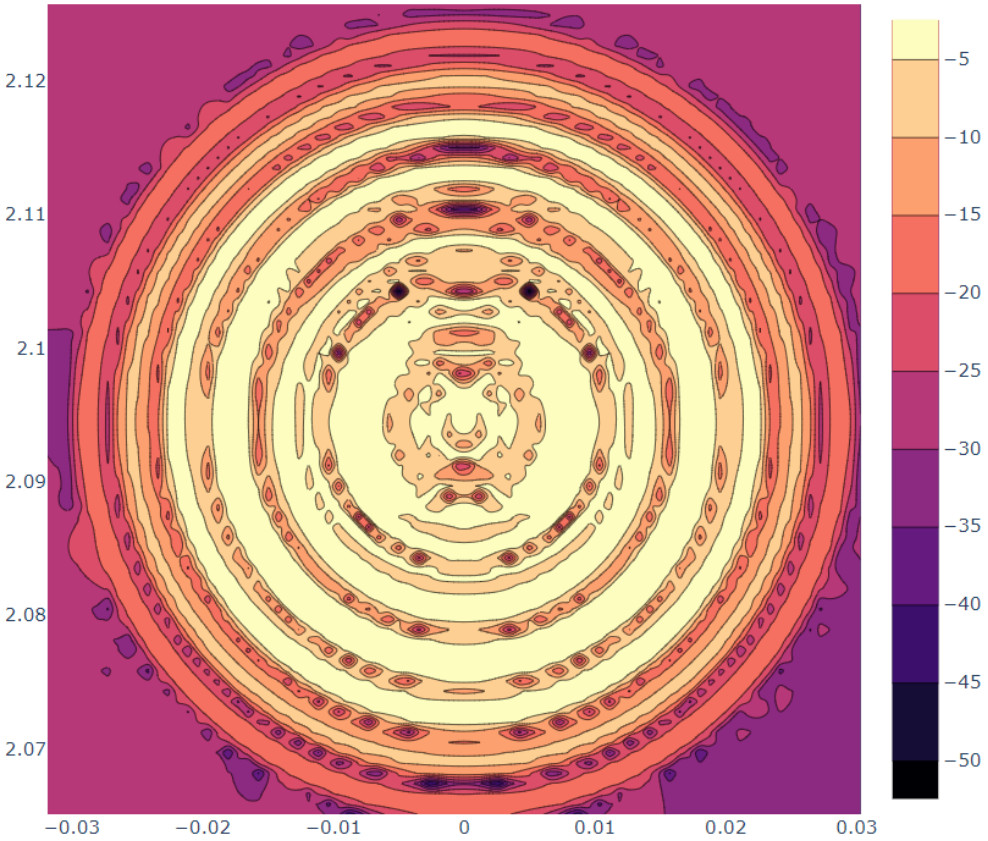


Рисунок 2 – Промежуточный кадр из анимации графиков типа density plot

надо определить четыре их компоненты:

$$j_1^{\text{cond}} = 8 \int \frac{d^2p}{(2\pi)^2} \frac{(p_1 - A_1(t_i))}{\varepsilon(p_1, p_2, t_i)} f_1(p_1, p_2, t_i). \tag{2}$$

$$j_2^{\text{cond}} = 8 \int \frac{d^2p}{(2\pi)^2} \frac{(p_2 - \frac{2\pi}{3} - A_2(t_i))}{\varepsilon(p_1, p_2, t_i)} f_1(p_1, p_2, t_i) \tag{3}$$

$$j_1^{\text{pol}} = -4 \int \frac{d^2p}{(2\pi)^2} \frac{(p_2 - \frac{2\pi}{3} - A_2(t_i))}{\varepsilon(p_1, p_2, t_i)} f_2(p_1, p_2, t_i) \tag{4}$$

$$j_2^{\text{pol}} = 4 \int \frac{d^2p}{(2\pi)^2} \frac{(p_1 - A_1(t_i))}{\varepsilon(p_1, p_2, t_i)} f_2(p_1, p_2, t_i) \tag{5}$$

Здесь первые две компоненты определяют ток проводимости, а последние две – поляризационный ток. Ток проводимости, как и плотность носителей заря-

да, определяется через функцию распределения $f(p_1, p_2, t)$, а поляризационный ток через вспомогательную функцию $u(p_1, p_2, t)$. Кроме того, в подынтегральных выражениях присутствуют значения координат $p_1, p_2$ и параметры внешнего электрического поля. В качестве последних выступают две компоненты векторного потенциала $A_1(t_i)$ и $A_2(t_i)$. Знаменатель

$$\epsilon(p_1, p_2, t_i) = \sqrt{(p_1 - A_1(t_i))^2 + \left(p_2 - \frac{2\pi}{3} - A_2(t_i)\right)^2} \tag{6}$$

тоже определяется через эти параметры. На рисунке 3 представлен график первой компоненты плотности тока проводимости.

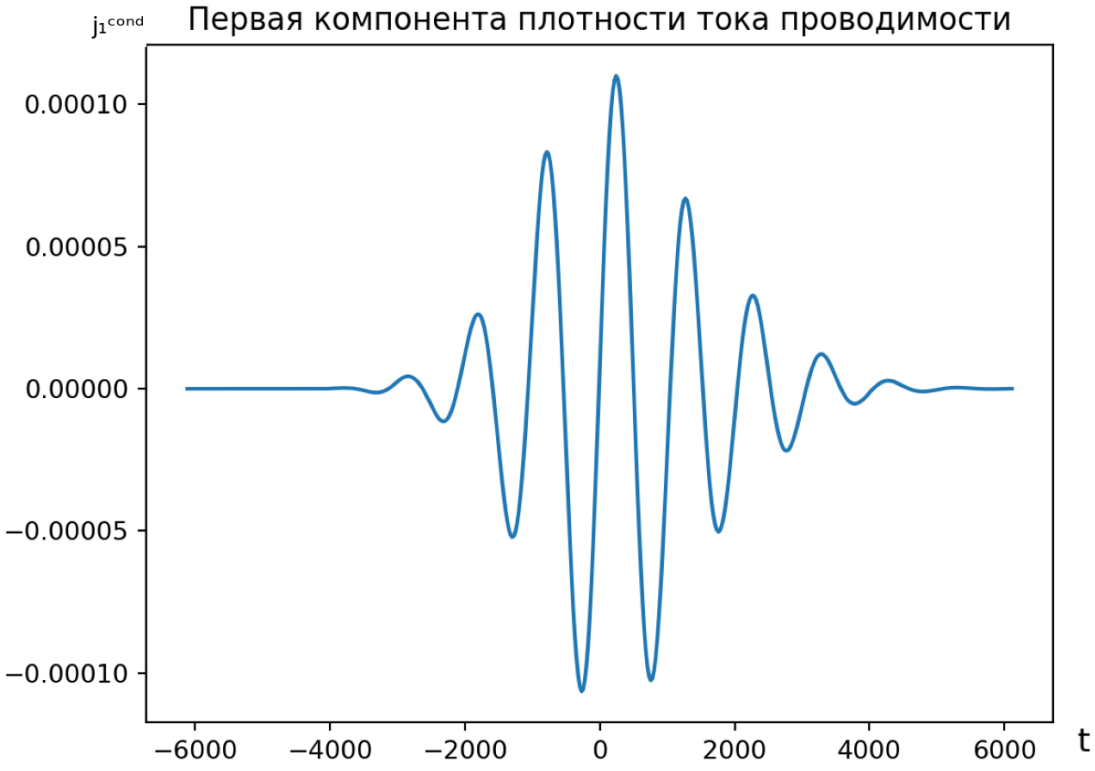


Рисунок 3 – График первой компоненты плотности тока проводимости

## ЗАКЛЮЧЕНИЕ

В ходе выполнения бакалаврской работы были изучены особенности работы с большими массивами данных различной природы. Основное внимание было сосредоточено на современных возможностях языка Python и специализированных библиотек на его основе. Изучены и проиллюстрированы различные способы визуального представления многомерных массивов, включая анимацию процессов, развивающихся во времени. На основе этого были решены две основные задачи, ставившиеся в начале работы: наглядная визуализация результатов численного моделирования отклика двумерной системы на возмущающего воздействия, причем средства анимации были использованы для этого впервые, и обработка первичных данных с уменьшением их размерности до одномерной временной последовательности наблюдаемых параметров. Последнее необходимо для сопоставления результатов моделирования с экспериментальными данными и дополнительного анализа наблюдаемых эффектов. Полученные результаты позволяют сделать вывод, что цель работы была успешно достигнута.

**Основные источники информации**:


1 Big Data Management and Processing / Kuan-Ching Li, Hai Jiang, Albert Y. Zomaya, 2017. – Chapman and Hall/CRC – 96 с. (Дата обращения: 05.03.2023) – Яз. англ.

2 Big Data: Challenges and Opportunities / S. Vinothina, S. Dhanalakshmi – International Journal of Advanced Research in Computer Science and Software Engineering, 2015. – 10 с. (Дата обращения: 10.03.2023) – Яз. англ.

3 How to Structure Modern Big Data Architecture? [Электронный ресурс] URL: https://nexocode.com/blog/posts/hadoop-spark-kafka-modern-big-data-architecture/ (Дата обращения: 14.03.2023) – Яз. англ.

4 Guide to NumPy / Travis E. Oliphant, 2006. – 55 с. (Дата обращения: 22.03.2023) – Яз. англ.

5 The Ultimate Guide to the NumPy Package for Scientific Computing in Python [Электронный ресурс] https://www.freecodecamp.org/news/the-ultimate-guide-to-the-numpy-scientific-computing-library-for-python/ (Дата обращения: 03.05.2023) – Яз. англ.

6 Pandas.DataFrame.query() by Examples [Электронный ресурс] https://sparkbyexamples.com/pandas/pandas-dataframe-query-examples/ (Дата обращения: 03.05.2023) – Яз. англ.

7 Visualizing Data in Python [Электронный ресурс]
URL: https://realpython.com/visualizing-python-plt-scatter/ (Дата обращения: 10.04.2023) – Яз. англ.
8 3D Surface plotting in Python using Matplotlib [Электронный ресурс] URL: https://www.geeksforgeeks.org/3d-surface-plotting-in-python-using-matplotlib/ (Дата обращения: 14.04.2023) – Яз. англ.