

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

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

Кафедра дискретной математики и информационных технологий

**АНАЛИЗ ЧАСТОТНЫХ ХАРАКТЕРИСТИК РЕЗУЛЬТАТОВ  
ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ**

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студента 4 курса 421 группы  
направления 09.03.01 — Информатика и вычислительная техника  
факультета КНиИТ  
Бисенгалиева Тимура Руслановича

Научный руководитель

доцент, к. ф.-м. н.

\_\_\_\_\_

А. Д. Панфёров

Заведующий кафедрой

доцент, к. ф.-м. н.

\_\_\_\_\_

Л. Б. Тяпаев

Саратов 2023

## ВВЕДЕНИЕ

Прогресс в развитии электронных вычислительных машин и телекоммуникационных систем привели к радикальным изменениям в сфере информационного обеспечения. Конечным потребителям теперь доступны практически неограниченные объёмы информации, которые далеко превосходят человеческие возможности по её обработке и анализу. Это сделало обязательным использование аппаратно-программных решений, берущих на себя наиболее трудоёмкую часть первичной обработки массивов данных в самых разных областях человеческой деятельности. Наиболее близкий и понятный большинству пример — системы поиска данных в Интернет. Ими мы сейчас пользуемся практически постоянно и уже не представляем других способов первичного поиска информации в глобальной сети. Но общедоступная сеть предоставляет доступ только к небольшой части накопленной информации. Огромными хранилищами данных располагают государственные структуры, корпорации, научные организации. Именно в науке был сформулирован термин «Большие данные» (Big Data) и разработаны первые методы и инструменты для их быстрой обработки и анализа. Сегодня владение соответствующими теорией и практикой работы в этой области является обязательным условием коммерческого успеха в бизнесе, прогресса в науке.

С примерами последнего типа можно познакомиться в области численного моделирования сложных физических систем, где генерация больших объёмов промежуточных данных требует применения к ним современных технологий обработки и анализа. Как правило, такая обработка и анализ выполняются поэтапно. На более поздних этапах глубина анализа и сложность применяемых методов увеличивается.

Целью представляемой работы является разработка и реализация программного решения для спектрального анализа временных рядов, получаемых после предварительной обработки первичных данных численных экспериментов по моделированию поведения отклика двумерной квантово механической системы на возмущающее воздействие. Для достижения поставленной цели необходимо решить следующие задачи:

- изучить особенности работы с большими данными;
- изучить характеристики и формы представления данных, с которыми необходимо работать;

- изучить методы спектрального анализа временных рядов;
- выбрать программные инструменты для чтения и подготовки данных;
- выбрать программные инструменты для определения спектра анализируемых временных рядов и представления полученных результатов;
- реализовать программное решение;
- протестировать его на предоставленных данных.

В данной работе содержится 4 главы, а именно «Анализ временных рядов: выявление закономерностей и прогнозирование в различных областях», «Big Data: возможности, проблемы и технологии для анализа огромных массивов данных», «Использование Python и математических библиотек для анализа результатов численного моделирования» и «Проведение анализа на реальных данных».

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе рассматривается понятие анализа временных рядов, спектрального анализа и его главного инструмента — преобразование Фурье. Временные ряды широко применяются в различных областях, таких как экономика, финансы, климатология, медицина, транспорт и многие другие. Они представляют собой последовательности данных, измеренных в разные моменты времени, и могут содержать ценные информационные закономерности, тренды и шумы. Понимание и анализ этих закономерностей является критическим для принятия обоснованных решений и прогнозирования будущих событий. Характеристики временного ряда зависят от того, являются ли данные неэкспериментальными или получены в результате какого-либо запланированного эксперимента. Так, временные ряды в экономике и социальных науках являются примерами неэкспериментальных данных. В технике и физических науках масштаб времени, в течение которого нужно собирать полезные данные, обычно намного меньше, так что можно получить временные ряды, содержащие гораздо больше значений [1].

Спектральный анализ является большой областью анализа временных рядов. Он позволяет выявлять различные спектральные характеристики временных рядов, которые содержат полезную информацию о сигнале. Например спектральная плотность мощности (СПМ) позволяет определить, какие частоты вносят наибольший вклад в сигнал и имеют наибольшую энергию.

Преобразование Фурье является основным инструментом спектрального анализа временных рядов. Оно позволяет разложить временной сигнал на составляющие синусоидальные волны различных частот. Преобразование Фурье преобразует временной сигнал из временной области в частотную область, представляя его спектр, который отображает амплитуды и фазы каждой составляющей частоты [2]. В случае дискретного временного ряда, преобразование Фурье применяется в форме дискретного преобразования Фурье (ДПФ). ДПФ является численным методом преобразования Фурье, который позволяет вычислить спектр временного ряда, рассчитывается по формуле:

$$X_k = \frac{1}{N} \sum_{i=0}^{N-1} x_n e^{-j \frac{2\pi ki}{N}} = \frac{1}{N} \sum_{i=0}^{N-1} x_n \left[ \cos \frac{2\pi ki}{N} - j \sin \frac{2\pi ki}{N} \right] \quad (1)$$

ДПФ для своей реализации требует выполнения  $N^2$  умножений ком-

плексных чисел, где  $N$  — количество элементов выборки. ДПФ может быть сильно упрощено, если использовать свойства симметрии и периодичности коэффициентов. Результатом переработки выражений для ДПФ является быстрое преобразование Фурье (БПФ), которое требует только  $\frac{N \log_2 n}{2}$  умножений комплексных чисел. Вычислительная эффективность БПФ по сравнению с ДПФ становится весьма существенной, когда количество точек БПФ увеличивается до нескольких тысяч. Очевидно, что БПФ вычисляет все компоненты выходного спектра. Если необходимо рассчитать только несколько точек спектра, ДПФ может оказаться более эффективным. Вычисление одного выходного отсчета спектра с использованием ДПФ требует только  $N$  умножений с комплексными числами.

**Во второй главе** рассматривается понятие Big Data, актуальность и проблемы технологии. Большие данные — это чрезвычайно большие и сложные массивы данных, которые невозможно легко управлять, обрабатывать или анализировать с помощью традиционных методов обработки данных [3]. Их можно описать как "3V": volume, velocity, variety (объем, скорость и разнообразие соответственно). Объем относится к огромному количеству данных, получаемых из различных источников, таких как социальные сети, датчики, транзакции и т. д. Скорость относится к скорости, с которой генерируются данные, которые необходимо обрабатывать в режиме реального времени или почти в режиме реального времени. Разнообразие относится к различным типам и форматам данных, включая структурированные, полуструктурированные и неструктурированные данные.

Большие данные обладают потенциалом для раскрытия ценной информации и более глубокого понимания сложных явлений. Их анализ позволяет выявить закономерности, корреляции и тенденции, что ведет к принятию решений на основе данных и получению действенных выводов. Данные находят применение во многих областях, включая бизнес, здравоохранение, финансы, социальные науки и многое другое [4].

Хотя большие данные предоставляют огромные возможности, они также создают значительные проблемы. Огромный объем и сложность данных требуют новых подходов и технологий для хранения, обработки и анализа. Проблемы конфиденциальности и безопасности возникают в связи с чувствительным характером собираемых данных. Кроме того, необходимо ре-

шить этические вопросы, касающиеся использования данных и потенциальной предвзятости.

**В третьей главе** для решения поставленной задачи анализа частотных характеристик результатов численного моделирования, было необходимо выбрать соответствующие инструменты. В качестве таковых было решено использовать Python и ряд математических библиотек, разработанных для него и обеспечивающих эффективную работу с массивами данных.

Python — это высокоуровневый язык программирования общего назначения, который стал одним из ведущих и популярнейших в сообществе программистов. По своим возможностям он классифицируется от разработки упрощенных приложений до проведения сложных математических вычислений с одинаковым уровнем сложности.

Numpy (Numerical Python) — это фундаментальная библиотека в экосистеме Python для научных вычислений и анализа данных. Она предоставляет инструменты и функции для эффективных численных операций и манипулирования массивами. С помощью объекта *ndarray* (N-мерный массив), который позволяет эффективно хранить и обрабатывать большие массивы данных, и векторных операций можно выполнять вычисления над целыми массивами одновременно.

SciPy — это научная библиотека для Python, предназначенная для математики, науки и инженерии. Она является открытым исходным кодом и распространяется по лицензии. SciPy основана на библиотеке NumPy, которая предоставляет удобные и быстрые возможности для работы с N-мерными массивами. Основная цель создания SciPy заключается в том, чтобы она работала с массивами NumPy и предоставляла множество эффективных и удобных численных методов, таких как рутинные процедуры для численного интегрирования и оптимизации. SciPy предоставляет функции для выполнения преобразования Фурье и обратного преобразования Фурье, что позволяет проводить спектральный анализ временных рядов. Функции *fft* и *rfft* в SciPy используются для выполнения БПФ [5].

Matplotlib — это популярная библиотека визуализации данных на языке Python, которая предоставляет широкий спектр инструментов для создания высококачественных графиков и визуализаций. Matplotlib может быть использован для визуализации и анализа больших наборов данных, позволяя

проникать в суть и эффективно передавать результаты.

Pandas — это библиотека манипулирования и анализа данных на языке Python, широко используемая в области Big Data. Pandas может эффективно работать с большими наборами данных, выполнять предварительную обработку данных и обеспечивать исследовательский анализ данных.

Jupyter Notebook — это веб-приложение с открытым исходным кодом, которое позволяет проводить интерактивные и исследовательские вычисления. Оно предоставляет среду, в которой пользователи могут создавать и совместно использовать документы, содержащие живой код, уравнения, визуализации и описательный текст.

Представленные выше функции были использованы для написания кода, обеспечивающие анализ частотных характеристик временных рядов. Для проверки его работоспособности были использованы два тестовых примера. В качестве таковых формировались две последовательности по 1000 отсчетов периодических функций:

$$f_1(t) = \sin(t) \quad (2)$$

$$f_2(t) = \sin(t) + 2\sin(2t) + 3\sin(3t) \quad (3)$$

Отсчеты брались на интервале значений  $t$  от 0 до  $10\pi$ , с постоянным шагом по времени составлявшим  $\frac{\pi}{100}$ . Далее оба полученных временных ряда подвергались преобразованию Фурье. В силу линейности преобразования Фурье, амплитуды фурье-образов имеют то же соотношение что и амплитуды соответствующих сигналов. При работе с реальными временными рядами соотношение амплитуд различных частотных компонент может быть очень велико. Поэтому на практике обычно используется их представление в логарифмическом масштабе, через преобразование к децибелам (дБ) по формуле  $10 \times \log |x|^2$ . Конечные результаты преобразований изображены на рисунке 1 (а,б).

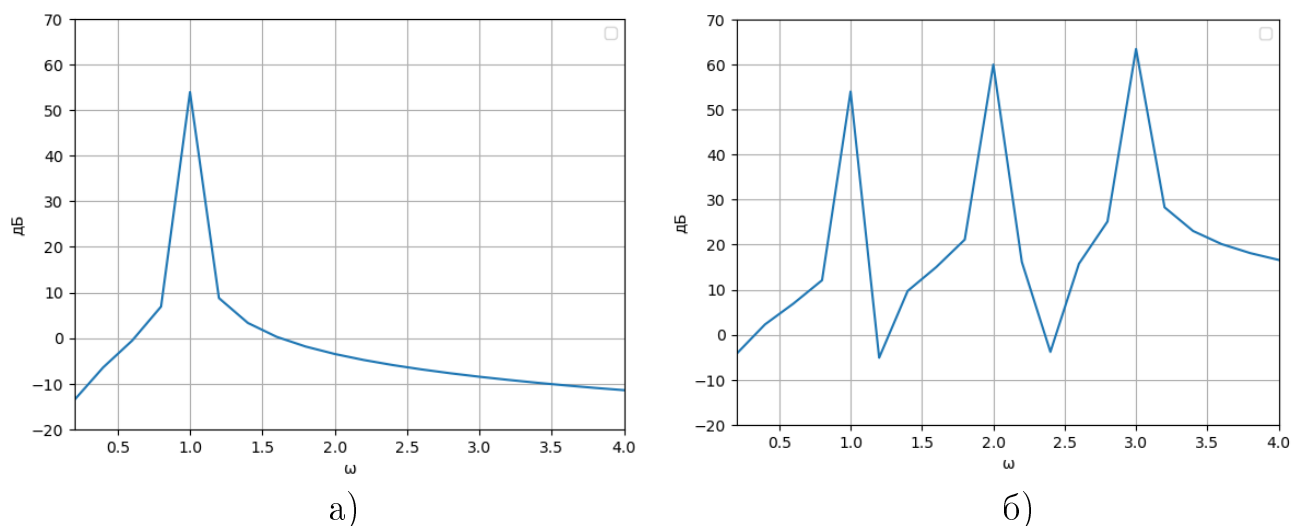


Рисунок 1 – Графики: а) Спектральная плотность мощности  $f_1$  в дБ, б) Спектральная плотность мощности  $f_2$  в дБ

**В четвертой главе** был проведен анализ на реальных данных. Данные для анализа были взяты из результатов численного моделирования эксперимента "Extremely effective terahertz high-harmonic generation in graphene by hot Dirac fermions" (Чрезвычайно эффективная терагерцовая генерация высоких гармоник в графене горячими фермионами Дирака) [6]. Он был проведен группой исследователей из Института Макса Планка по изучению структуры и динамики материи в Гамбурге, Германия. Целью эксперимента было исследование генерации высокочастотных гармоник в графене с помощью терагерцовых (ТГц) импульсов.

Генерация множественных оптических гармоник — умножение энергии фотона в результате нелинейного взаимодействия между светом и материей — является ключевой технологией в современной электронике и оптоэлектронике, поскольку она позволяет преобразовывать оптические или электронные сигналы в сигналы с гораздо более высокой частотой и генерировать частотные гребенки. Благодаря уникальным электронным свойствам графена, для описания которых используются безмассовые фермионы Дирака, было неоднократно предсказано, что генерация оптических гармоник в графене должна быть особенно эффективной на технологически важных терагерцовых частотах. С момента публикации данной работы эти предположения были неоднократно подтверждены экспериментально и обоснованы теоретически. Для моделирования процессов такого типа в СГУ им. Чернышевского был разра-



ботан отдельный подход.

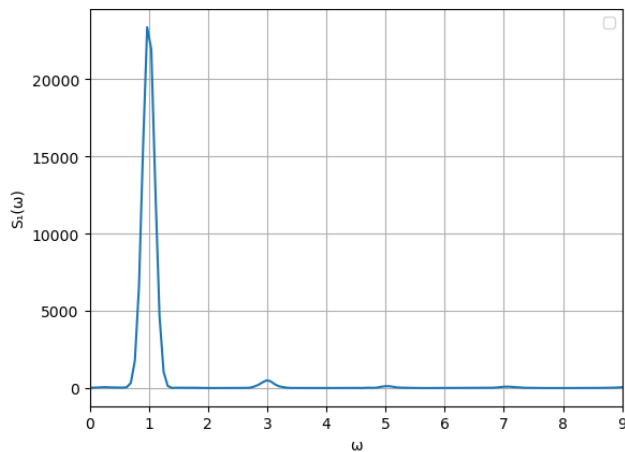
Если применить преобразование Фурье к результатам численного моделирования эксперимента, то можно получить спектр высокочастотных гармоник, генерируемый образцом графена. Частотный спектр, полученный в результате преобразования Фурье, предоставит ценную информацию о свойствах высокочастотных гармоник, таких как его частотные компоненты и интенсивность. Эта информация может быть использована для дальнейшего анализа высокочастотных гармоник и лучшего понимания основных физических процессов, которые его порождают.

Для анализа были взяты результаты моделирования процессов, протекающих в графене при воздействии на него короткого импульса с частотой  $6.3 \times 10^{13}$  Гц. Они представляют собой одну компоненту поверхностной плотности тока проводимости и одну компоненту поверхностной плотности поляризационного тока. Эти данные представлены в виде временных рядов содержащих по 1081 отсчету, разнесенному во времени на  $2.116 \times 10^{-16}$  с. Общий интервал времени составляет  $2.285 \times 10^{-13}$  с.

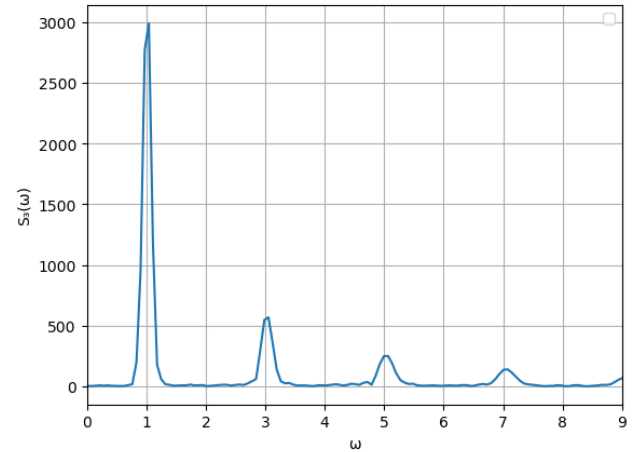
Результаты Фурье преобразования временных рядов приведены на рисунке 2 (а,б). Горизонтальная ось представляет из себя шкалу частот, где за единицу принимается частота внешнего возмущающего импульса  $6.3 \times 10^{13}$  Гц. Рисунки демонстрируют присутствие основной компоненты, соответствующей частоте действующего сигнала и ряду нечетных гармоник, с третьей по девятой. Высота соответствующих пиков на рисунке 2 (а), но они явно видны на рисунке 2 (б).

На рисунке 3 (а,б) представлено распределение по частоте СПМ сигналов в дБ. Они дают наглядную картину распределения высокочастотных гармоник в спектре анализируемых временных рядов.

Представленные распределения хорошо коррелируют с результатами представленными в работе "Extremely effective terahertz high-harmonic generation in graphene by hot Dirac fermions"[6].

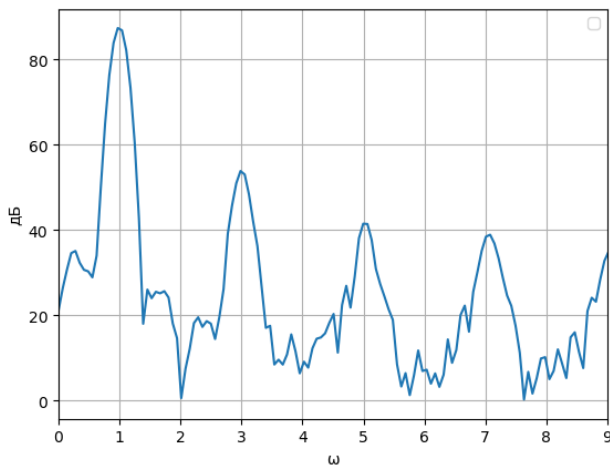


а)

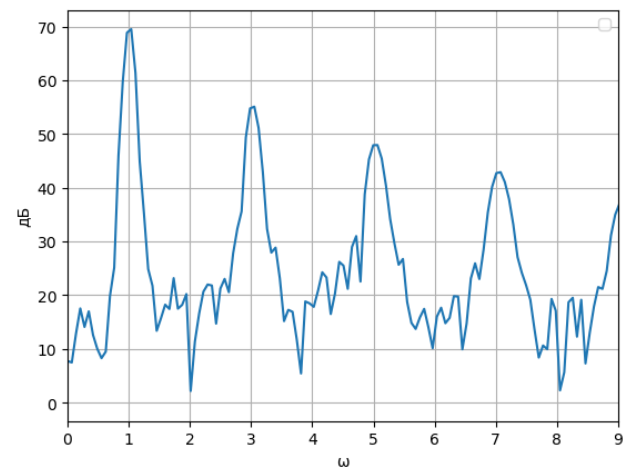


б)

Рисунок 2 – Графики: а) Амплитуда Фурье-образа  $j_1^{cond}(t)$ , б) Амплитуда Фурье-образа  $j_1^{pol}(t)$



а)



б)

Рисунок 3 – Графики: а) Спектральная плотность мощности  $j_1^{cond}(t)$  в дБ, б) Спектральная плотность мощности  $j_1^{pol}(t)$  в дБ

## ЗАКЛЮЧЕНИЕ

В результате проделанной работы было реализовано и апробировано программное решение для спектрального анализа временных рядов, получаемых после предварительной обработки первичных данных численных экспериментов по моделированию поведения отклика двумерной квантово механической системы на возмущающее воздействие.

Были изучены особенности работы с большими массивами данных и используемые для этого инструменты. Основной акцент делался на анализе дискретных временных рядов в соответствии с целью выполнявшейся работы. Изучались как теоретические основы такого анализа, так и имеющиеся инструменты. Сделанный выбор позволил продемонстрировать выполнение анализа спектрального состава модельных гармонических сигналов и показать результативность разработанного решения применительно к результатам моделирования поведения реалистичной физической системы на основе графена в которой имеет место нелинейный процесс генерации высокочастотных гармоник.

Поставленные задачи были полностью решены. Цель работы успешно достигнута.

### **Основные источники информации:**

- 1 Афанасьев В.Н Анализ временных рядов и прогнозирования / Афанасьев В.Н., Юзбашев М.М. – М.: Финансы и статистика, 2001. 228 с.
- 2 Витязев В.В. Спектрально-корреляционный анализ равномерных временных рядов / Витязев В.В. – СПб.: Изд-во С.-Петербур. ун-та, 2001. 48 с.
- 3 What is big data analytics?: [Электронный ресурс] URL: <https://www.techtarget.com/searchbusinessanalytics/definition/big-dataanalytics> (Дата обращения: 21.04.2023) – Яз. англ.
- 4 Mayer-Schönberger V. Big Data: A Revolution That Will Transform How We Live, Work, and Think / Mayer-Schönberger V., Cukier K. – Harper Business, 2014. 272 с. – Яз. англ.
- 5 Fourier Transforms With scipy.fft: Python Signal Processing: [Электронный ресурс] URL: <https://realpython.com/python-scipy-fft/> (Дата обращения: 10.04.2023) – Яз. англ
- 6 Extremely efficient terahertz high-harmonic generation in graphene by hot

Dirac fermions: [Электронный ресурс] URL: <https://www.nature.com/articles/s41586-018-0508-1> (Дата обращения: 20.03.2023) – Яз. англ.