

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 ДАННЫХ В МНОГОМЕРНЫХ МАССИВАХ И ИХ
ВИЗУАЛИЗАЦИЯ ПРИ ИССЛЕДОВАНИИ ЭФФЕКТОВ В
ЭКСТРЕМАЛЬНО СИЛЬНЫХ ЭЛЕКТРИЧЕСКИХ ПОЛЯХ**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 271 группы

направления 09.04.01 «Информатика и вычислительная техника»

факультета компьютерных наук и информационных технологий

Аль Карави Раада Джоудах Салмана

Научный руководитель

к. ф.-м.н., доцент

А.Д. Панферов

подпись, дата

Зав. кафедрой

к. ф.-м.н., доцент

Л.Б. Тяпаев

подпись, дата

Саратов 2019

ВЕДЕНИЕ

Современные высокопроизводительные вычислительные комплексы позволяют выполнять имитационное моделирование для различных типов систем, объектов и процессов. Значительная часть задач моделирования требует в общем случае оперировать полевыми характеристиками систем и процессов размерностью не менее трёх. Использование многомерных моделей не только усложняет задачу с вычислительной точки зрения, но и делает нетривиальной процедуру интерпретации результатов моделирования.

Задачей работы было обеспечить наглядное и информативное представление результатов моделирования. Размерность анализируемых и представляемых данных может достигать четырех. В качестве одной из рассматриваемых координат выступает время. Представленные результаты позволяют рассматривать весь процесс эволюции таких систем от включения внешнего электрического поля, до его выключения или любого промежуточного момента.

Для достижения поставленной цели было необходимо решить следующие задачи:

1. Изучить технологические особенности работы с большими многомерными массивами данных, основные способы анализа и визуализации данных;
2. Изучить характер и форму представления данных, предоставляемых для обработки, процедуры анализа данных, обеспечивающие их правильную интерпретацию;
3. Реализовать процедуры обработки и анализа данных и продемонстрировать их на реальной модели.

Магистерская работа состоит из введения, 4 разделов, заключения и списка использованных источников. Общий объём работы составляет 53 страницы, включая 24 рисунка. Список использованных источников включает 20 наименований.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Массивы данных» посвящен определению основных характеристик и принципов организации массивов данных, используемых в информационных технологиях для описания различных процессов и явлений. Приведено определение понятия «массив» и «многомерный массив», способы их реализации, индексирования. Дано определение «структуры» данных, как обобщение понятия массива. Определены понятия плотных и разреженных массивов, регулярности размерности массива. Как важные типы массивов рассмотрены векторы и матрицы. Применение массивов различной размерности проиллюстрировано на примере работы с географическими информационными системами.

Работа с массивами достаточно стандартизована в современных языках программирования. Наиболее удобными для использования являются массивы с фиксированным размером или с размером, максимальное значение которого известно на момент написания программы. Более сложной является ситуация, когда размер массива заранее не известен. В этом случае возможны две альтернативы: размер массива заранее не известен, но он не меняется в ходе выполнения программы, и наиболее сложный вариант, когда размер массива может меняться в ходе выполнения программы. Варианты организации работы с массивами с использованием статического и динамического выделения памяти рассмотрены.

Рассматриваются также специализированные инструменты и библиотеки для работы с массивами данных. В качестве примеров приводится стандартная библиотека шаблонов для C++ и библиотека Thrust, используемая при написании программ обработки для графических сопроцессоров.

Отдельно определено понятие анализа данных в контексте выполняемой работы. Продемонстрированы простейшие варианты визуализации на примере демонстрации трех различных функциональных завистей от одного параметра. Для этого использовались данные из предметной области работы.

Второй раздел «Функция распределения носителей в задачах о поведении квантово-полевого вакуума во внешнем электрическом поле» даёт развернутое представление о процессах, данные по моделированию которых являются предметом обработки, анализа и визуализации. Это достаточно сложные и трудно наблюдаемые явления, предсказываемые и описываемые современной квантовой электродинамикой. Они лежат далеко за пределами обыденного человеческого опыта на границе возможностей самых мощных экспериментальных лазерных установок. В силу сложности соответствующей теории, в данной части изложение носит описательный характер.

Более детально изложена математическая модель, являющаяся непосредственным генератором данных. Это система трех дифференциальных уравнений с переменными коэффициентами, определяющая взаимосвязанную эволюцию трех функций. Ключевую роль для последующей обработки играет первая из них, имеющая смысл функции распределения электронов, генерируемых электрическим полем. Хотя само решение этой системы уравнений не является предметом представляемой работы, важно понимать взаимосвязь обрабатываемых параметров и смысл изображений, получаемых при их визуализации.

Здесь же даны определения наблюдаемых интегральных характеристик:

$$n(t) = 2 \int \frac{d^3p}{(2\pi)^3} f(\vec{p}, t),$$

$$\vec{j}_c(t) = 2e \int \frac{d^3p}{(2\pi)^3} \frac{\vec{p} - e\vec{A}}{\varepsilon(\vec{p}, t)} f(\vec{p}, t),$$

вычисление которых является элементом предстоящего анализа данных. Эти величины определяют плотность числа рожденных частиц и генерируемую ими плотность тока. Они потенциально наблюдаемы в эксперименте и их оценка по результатам моделирования является важной целью. Их получение в форме временных рядов позволяет искать корреляции между ними, между ними и поведением функции распределения в полном пространстве аргументов и,

в перспективе, особенности зависимости от времени этих интегральных параметров.

Третий раздел «Обработка данных» включает полное описание использовавшихся для обработки и анализа данных инструментов. В его начале приведены характеристики моделирующих программ, аппаратных платформ на которых они работают и форматов вывода и сохранения данных.

Поскольку среди используемых вариантов моделирующих программ есть параллельные решения на платформе системы компьютерной алгебры Mathematica, было принято решение использовать эту платформу и для реализации процедур обработки и анализа результатов.

Для работы с массивами данных в пакете Mathematica используется встроенные возможности, обеспечивающие высокоуровневые операции с объектами такого типа. Массивы могут формироваться непосредственно с использованием процедуры Array, от простейшего варианта Array[f,n], который создаёт линейный массив размером n с использованием определенной предварительно функции f вида {f[1], f[2], f[n-1], f[n]}, и до вариантов вида Array[f, {n1, n2, n3, ...}, {{a1, b1}, {a2, b2}, {a3, b3,}}]. В последнем случае генерируется многомерный массив. Для каждой размерности определяются n_i значений аргумента, которые будут взяты из интервала $\{a_i, b_i\}$.

Главным элементом форматирования массива является список (List), оформляемый как перечень однотипных элементов, разделяемых запятыми и заключенных в фигурные скобки вида {a, b, c, d, e, f}. Многомерный массив оформляется как список вложенных списков, например в двумерном случае это будет {{a, b, c}, {d, e, f}, {i, j, k}}.

При необходимости мы можем обратиться к любому вложенному списку или даже отдельному элементу на нижнем уровне вложения. Это реализуется процедурой вида Extract[L, n]. Доступ к глубоко вложенным элементам осуществляется последовательным использованием этой процедуры вида Extract[Extract[Extract[L, l], m], n].

Процедуры переформатирования исходных массивов, формирования отдельных их сечений, формирование массивов меньшей размерности для визуализации требуют просмотра и выборки необходимых данных. Они реализуются с использованием циклов и последующей пересборки нового массива. Приведем пример такой процедуры:

```
Do[ mmm = AppendTo[mmm, {Extract[
  Extract[Extract[nnn, IntegerPart[k1]], IntegerPart[k2]], 1],
  Extract[Extract[Extract[fbn, IntegerPart[k1]], IntegerPart[k2]],
  2], Extract[Extract[Extract[fbn, IntegerPart[k1]], IntegerPart[k2]],
  4]}], {k1, 1, m}, {k2, 1, n]
```

Эта процедура из двумерного массива размером $m \times n$, элементами которого являются строки из нескольких (не менее четырех) элементов, формирует одномерный массив строк длиной $m * n$. В новом массиве каждая строка содержит только три элемента, занимавших позиции **1, 2, 4** в строках исходного массива.

Приведены характеристики и варианты применений функций **ListPlot[...]**, **ListPlot3D[...]** и **ListDensityPlot[...]**, предоставляющих разнообразные возможности визуализации линейных и двумерных массивов. Их работа проиллюстрирована на простейших примерах.

Численное интегрирование реализовано кратным вложенным суммированием по элементам с фиксированным значением времени. В качестве примера приведем простейший интеграл в трехмерном Декартовом импульсном пространстве:

$$n(t_i) = 2/2\pi^3 \text{Sum}[\Delta p_1 \Delta p_2 \Delta p_3 f[j, k, l, i, 4], \{j, 1, n1\}, \{k, 1, n2\}, \{l, 1, n3\}]$$

Предполагается, что каждой точке t_i на временной оси соответствует трехмерный массив значений функции распределения в импульсном пространстве размером $n1 \times n2 \times n3$. Интегрирование для компонент плотности тока выполняется аналогично.

Четвертый раздел «Представление результатов» демонстрирует использование разработанных процедур для обработки, анализа и визуализации данных, полученных в ходе моделирования поведения КЭД – вакуума в поле лазерного импульса вида (3) с частотой, соответствующей энергии квантов в 0.1 от энергии массы покоя электрона, амплитудой поля в 0.2 от критического значения $E_{cr} = m^2/e = 1.323 * 10^{18}$ В/м и характерной длительностью в 12 периодов основной частоты.

Данные передавались в форме текстового файла, содержащего трехмерный массив строковых элементов вида:

{0., -1.5, -314.159, 1.96144*10^-8, 2.05493*10^-6}

Каждая строка включает пять параметров. Первые три из которых являются координатами. В случае поля с постоянным направлением результаты моделирования имеют в импульсном пространстве осевую симметрию. В этом случае удобно воспользоваться цилиндрической системой координат, в которой результаты моделирования не зависят от текущего значения полярного угла, а только от расстояния от оси \overline{OZ} и значения z (p_3 – в выбранных ранее обозначениях). С учетом этого данные включают только значения p_1 – первое в списке, интерпретируется как расстояние от оси \overline{OZ} при нулевом значении полярного угла, и p_3 – второе в списке. Третье значение в списке является текущим значением времени.

Четвертый и пятый параметры представляют из себя собственно значение функции распределения $f(\vec{p}, t)$ для данной пространственно - временной точки и значение вспомогательной функции $u(\vec{p}, t)$. Значение вспомогательной функции $v(\vec{p}, t)$ не включалось в обрабатываемые данные.

Размер области импульсного пространства, для которой воспроизводилось поведение моделируемой системы и в пределах которой менялись значения импульсных координат составлял: $0 \leq p_1 \leq 1.5$, $-1.5 \leq p_2 \leq 1.5$. Моделирование выполнялось на сетках с различным шагом от 0.25 до 0.01. Интервал времени, для которого записывались данные, охватывал только центральную часть импульса, представленного на рисунке 15: $-314.159 \leq t \leq$

314.159. Дискретизация этого интервала выполнялась с постоянным шагом с разбиением на 1000 равных частей. Всего получалось 1001 точка на временной оси, для которых записывались данные.

На первом этапе выполнялась визуализация функции распределения в момент времени максимального значения напряженности поля. В работе приведены соответствующие сечения, полученные на разных сетках и проведено их сравнение. Аналогичная визуализация выполнена для момента времени, когда эффективная амплитуда внешнего поля составляет всего 0.032, т.е. поле почти выключено.

Переход к интегральным характеристикам выполнялся с использованием двумерного обобщения метода средних прямоугольников. Результат оформлялся в виде промежуточного линейного массива, состоящего из элементов вида $\{t_i, n_i, j_{ci}\}$. Первый элемент определяет момент времени, второй значение плотности квазичастиц для этого момента времени и третий элемент – это плотность тока для данного момента времени.

Наличие такой компактной информации о поведении моделировавшейся системы позволило провести сопоставление поведения временных рядов n_i, j_{ci} , особенно наглядное при наложении этих графиков на одном рисунке.

На последнем этапе вновь выполнялось построение сечений функции распределения. Но теперь значения времени для построения сечения выбирались по результатам анализа временных рядов для интегральных характеристик. При выборе моментов времени, привязанных к минимальным значениям плотности числа частиц было обнаружено, что в это время функция распределения хотя и имеет относительно небольшие значения, но не обращается в ноль. Кроме того было обнаружено, что максимум функции распределения перемещается по импульсному пространству с частотой действующего поля. Процедура визуализации позволяет продемонстрировать это явно и вполне наглядно. Все перечисленные результаты подробно проиллюстрированы.

ЗАКЛЮЧЕНИЕ

В ходе выполнения выпускной квалификационной работы были изучены особенности обработки, представления и визуализации многомерных массивов данных. В настоящее время для практической информатики работа с большими массивами данных в разных конкретных областях и реализациях очень актуальна. В том числе с многомерными регулярными и не регулярными массивами данных приходится сталкиваться во многих технических и научных задачах.

В качестве обрабатываемых данных в работе использовались массивы, генерируемые при исследовании модели поведения физического вакуума в экстремально сильных электрических полях, построенной на основе квантового кинетического уравнения. Такой подход позволяет получать точное и адекватное описание процессов, представляющих большой интерес в связи с наращиванием мощности экспериментальных лазерных установок и приближения их параметров к значениям, при которых упомянутые процессы и эффекты будут играть существенную роль. Но промежуточные данные, получаемые в процессе такого моделирования, чрезвычайно объёмны и в определенном смысле избыточны. Для интерпретации и оценки получаемых результатов необходима их предварительная обработка и анализ.

Поставленная задача была выполнена и был разработан набор инструментов для обработки, анализа и визуализации таких данных. Он позволяет выполнять предварительную обработку и визуализацию заданных временных сечений исходных данных. Позволяет вычислять интегральные обобщающие характеристики и представлять их зависимость от времени в форме, удобной для последующего изучения и анализа.

Полученные инструменты были успешно опробованы на предоставленных данных, полученных в ходе моделирования реалистических полевых конфигураций, и позволили обеспечить их успешный анализ и интерпретацию.

Основные источники информации:

- 1 WolframAlpha [Электронный ресурс]: - URL <http://www.wolfram.com/mathematica/> (дата обращения: 12.03.2019). - Загл. с экрана. - Англ. рус.
- 2 С.А.Смолянский, Соросовский образовательный журнал, том 7, вып. 2, Москва,2001-69стр.
- 3 J.S. Schwinger, Phys. Rev. 82, БХВ СПб,- 2001г-600стр.
- 4 Журнал «Элементы» [Электронный ресурс]: - URL https://elementy.ru/novosti_nauki/433343/Nobelevskaya_premiya_po_fizike_2018 (дата обращения: 25.01.2019). - Загл. с экрана. - Яз. рус.
- 5 S.M. Schmidt, D. Blaschke, G. Reopke, S.A. Smolyansky, A.V. Prozorkevich, and V.D. Toneev, Int. J. Mod. Phys. E 7, 2008г-709стр
- 6 D.B. Blaschke, S.A. Smolyansky, A.D. Panferov, and L. Juchnowski, Proceedings of the Helmholtz International Summer School 2016 (HQ 2016) July 18-30, 2016 BLTP, JINR, Dubna, Russia DESY-PROC-2016-04 1 2017.
- 7 S.A. Smolyansky, A.D. Panferov, D.B. Blaschke, and N.T. Gevorgyan, Particles 2, p. 208 -г2019.
- 8 Microsoft server [Электронный ресурс]: - URL https://www.mssoft.ru/Makers/Microsoft/Windows_Server_2008_R2_Enterprise/ (дата обращения: 5.01.2019). - Загл. с экрана. - Яз. рус.
- 9 ПРЦ НИТ. Регламент доступа к кластеру [Электронный ресурс]: - URL http://prcnit.sgu.ru/sites/default/files/files/reglament_dostupa_%20k_klasteru_vysoko_proizvoditelnyh_vychisleniy_prc_nit.pdf (дата обращения: 15.01.2019). - Загл. с экрана. - Яз. рус.
- 10 GNU PLOT [Электронный ресурс]: - URL <https://www.gnu.org/software/gsl/doc/html/index.html> (дата обращения: 15.12.2018). - Загл. с экрана. - Яз. рус.