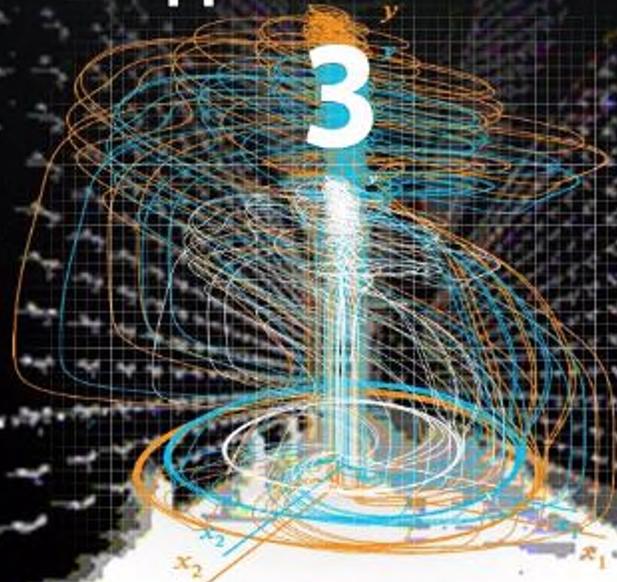


**ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ
В ЕСТЕСТВЕННЫХ НАУКАХ**

**МЕТОДЫ
СУПЕРКОМПЬЮТЕРНОГО
МОДЕЛИРОВАНИЯ**

3



СБОРНИК ТРУДОВ

МЕХАНИКА, УПРАВЛЕНИЕ И ИНФОРМАТИКА

**МОСКВА
2015**

ИНСТИТУТ КОСМИЧЕСКИХ ИССЛЕДОВАНИЙ
РОССИЙСКОЙ АКАДЕМИИ НАУК

ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ В ЕСТЕСТВЕННЫХ НАУКАХ

МЕТОДЫ СУПЕРКОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

Часть 3

СБОРНИК ТРУДОВ

17–19 ноября 2015 года, Россия, Таруса

Под редакцией
Р.Р. Назирова, Л.Н. Щура

Серия «Механика, управление и информатика»

МОСКВА
ИКИ РАН
2015

**Вычислительные технологии в естественных науках.
Методы суперкомпьютерного моделирования
Сборник трудов**

Россия, Таруса, 17–19 ноября 2015 г. Под ред. Р. Р. Назирова, Л. Н. Шура

В сборник вошли статьи, основанные на докладах, представленных на конференции «Методы суперкомпьютерного моделирования», которая проходила 17–19 ноября 2015 г. в Тарусе Калужской области на базе гостиницы «Интеркосмос» Института космических исследований Российской академии наук (ИКИ РАН). Это мероприятие уже третье в серии расширенных конференций и семинаров, где делается акцент на методах суперкомпьютерного моделирования и их применении в естественных науках. Мероприятия проводятся с 2009 г., на первом из них рассматривались вопросы компьютерного моделирования актуальных задач физики и механики. Второй семинар был посвящён обсуждению физических и медицинских аспектов исследования сложных нелинейных физических процессов в организме человека. На третьем обсуждались проблемы компьютерного моделирования задач биологии и химии. На четвёртом — проблемы технического зрения. На пятом — вопросы моделирования событий глобальных масштабов. Шестой семинар был посвящён исследованиям в области перспективных компьютерных систем, с акцентом на обсуждение концептуальных подходов, разработке новых методов и устройств. На седьмом семинаре рассматривались алгоритмы и методы в области вычислительной физики, причём особый акцент делался на обсуждении новых физических результатов. Материалы этого семинара можно найти на страницах семинара <http://www.iki.rssi.ru/seminar>. Восьмой получил название «Инфокоммуникационные технологии в научных исследованиях». Девятое и десятое мероприятия, как и нынешнее, одиннадцатое, получили название «Методы суперкомпьютерного моделирования», что связано с поддержкой семинаров с такой тематикой в рамках выполнения научных исследований по проекту РФФ № 14-21-00158 «Разработка алгоритмов и методов для задач математического моделирования на суперкомпьютерных системах, включая гибридные».

Ключевые слова: математическое моделирование, гибридные компьютерные системы, моделирование на суперкомпьютерах, методы и алгоритмы параллельных вычислений.

**Computer Technologies in Sciences.
Methods of Simulations on Supercomputers
Proceedings**

Russia, Tarusa, November 17–19, 2015. Eds. R. R. Nazirov, L. N. Shchur

This issue is the collection of the papers based on the talks presented at the extended workshop “Methods of simulations on supercomputers”, which was held at 17–19 November 2014, in Tarusa hotel Interkosmos of the RAS Space Research Institute. It is the Conference in the series of extended workshops and conferences, devoted to the computer technologies in natural sciences. The first workshop was devoted to computer simulations of the actual problems in physics and mechanics and papers was published in the issue 1, the second workshop was devoted to the discussion of the physical and medical aspects of the research of the complex non-linear processes in the human body, the third workshop was devoted to the computer simulations in biology and chemistry, the fourth was devoted to the discussion of the computer vision in control systems of the mobile objects. The fifth workshop was devoted to the simulation of events of global scale. Sixth workshop was devoted to the investigations of promising computing systems, with emphasis on the conceptual approaches and on the development of new methods and tools. Algorithms and methods in computational physics was discussed at the seventh workshop with the special accents on the physical results. Eighth workshop was transformed into the conference “Infocommunication technologies in scientific research” with the papers published in the double volume issue of journal Information Society. Ninth and tenth scientific events, as the current one, named “Methods of simulations on supercomputers”. The name is connected with the fact that workshops are supported within the research program the grant by Russian Scientific Foundation No. 14-21-00158 “Algorithms and methods for mathematical simulations on supercomputing systems, including hybrid ones”..

Keywords: simulations, hybrid computing systems, simulations on supercomputers, methods and algorithms of parallel computing.

Руководитель издания: *Зайцев Ю. И.*
Редакторы: *Полякова Т. В., Зайцев Ю. И., Виноградова С. Е.*
Компьютерная вёрстка: *Комарова Н. Ю.*
Дизайн обложки: *Захаров А. Н.*

Мнение редакции не всегда совпадает с точкой зрения авторов статей.
Отдельные статьи даны в авторской редакции.
Качество иллюстраций соответствует предоставленному авторами материалу.

Электронная версия сборника размещена на сайтах ИКИ РАН <http://www.iki.rssi.ru/print.htm>
и Российской научной электронной библиотеки <http://elibrary.ru/>.

СОДЕРЖАНИЕ

Методы суперкомпьютерного моделирования (Вместо предисловия)	5
<i>Алешин С. В.</i> Качественная структура процесса распространения возмущений в уравнении Колмогорова-Петровского-Пискунова с отклонением по пространственной переменной	9
<i>Биряльцев Е. В., Галимов М. Р., Демидов Д. Е., Елизаров А. М.</i> Опыт разработки и эксплуатации суперкомпьютерного комплекса для решения обратных задач сейсморазведки.	18
<i>Богданов А. В., Дегтярев А. Б., Храмушин В. Н.</i> Трёхмерная тензорная математика вычислительных экспериментов в гидромеханике	34
<i>Воробьёв А. Э., Титов А. Ю., Гаврилин В. А., Меньшиутин А. Ю., Бахирев И. А.</i> Транспортная модель Московского региона.	49
<i>Глызин С. Д., Колесов А. Ю., Розов Н. Х.</i> Неклассические релаксационные колебания и бифуркация голубого неба в задаче моделирования кардиоритмов	63
<i>Глызин С. Д., Колесов А. Ю., Розов Н. Х.</i> Режимы двухкластерной синхронизации в полносвязной сети релаксационных осцилляторов	82
<i>Горшков О. Н., Дербенева Н. В., Мухаматчин К. Р., Сатанин А. М.</i> Численное моделирование электронных состояний в диоксиде циркония с кислородными вакансиями и примесями иттрия	96
<i>Дербенева Н. В.</i> Влияние фосфора на Оже-рекомбинацию в нанокристаллах кремния	107
<i>Дружинин О. А., Троицкая Ю. И.</i> Прямое численное моделирование двухфазного турбулентного пограничного слоя над взволнованной водной поверхностью	114
<i>Ивановский Л. И.</i> Динамические свойства одного класса импульсных систем.	126
<i>Коноплев В. В., Боярский М. Н., Назиров Р. Р.</i> SCARP: масштабируемая и отказоустойчивая платформа лёгкой виртуализации для интенсивной работы с большими данными и облачных приложений.	132

<i>Леоненков С. Н., Жуматий С. А.</i> Оптимизация алгоритма Backfill и системы планирования заданий для использования на суперкомпьютере «Ломоносов»	140
<i>Попов И. С., Прудников П. В., Прудников В. В.</i> Численное описание неравновесного критического поведения двумерной структурно неупорядоченной XY-модели методами Монте-Карло	150
<i>Рыбакин Б. П.</i> Математическое моделирование процессов фрагментации и абляции молекулярного облака сильной ударной волной	161
<i>Хохлов Н. И., Петров И. Б.</i> Применение современных высокопроизводительных технологий для моделирования задач сейсмологии и геофизики	173
<i>Шур Л. Н.</i> К вопросу об информационно-телекоммуникационной инфраструктуре, а также о возможности проведения учреждениями ФАНО исследований в области интенсивной обработки данных с целью извлечения нового знания.	186

МЕТОДЫ СУПЕРКОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ (ВМЕСТО ПРЕДИСЛОВИЯ)

Читателю предлагаются избранные статьи, написанные на основе приглашенных докладов на Конференции «Методы суперкомпьютерного моделирования», прошедшей 17–19 ноября 2015 г. на базе гостиницы «Интеркосмос» Института космических исследований Российской академии наук. Как и предыдущие два мероприятия, конференция была посвящена методам и практике разработки и применения методов моделирования на суперкомпьютерах для решения задач в различных областях естественных наук, а также вопросам применения новых информационных и телекоммуникационных технологий при проведении исследований в области естественных наук. Особое внимание уделялось новым разработкам, идеям и методам, а также ярким результатам, добиться которых без применения современных методов проведения моделирования и без использования новых информационных и телекоммуникационных технологий было бы затруднительно или даже невозможно.

Особенность проводимых нами научных мероприятий [1–9] состоит в активном участии в них молодых научных сотрудников и аспирантов, число которых составило примерно половину докладчиков. Другая особенность — в их рамках проводились научно-учебные семинары с приглашёнными студентами, которые делают первые шаги в научных исследованиях по тематике семинара. Ещё одна особенность научных мероприятий. Начиная с девятой конференции их проведение было поддержано проектом РНФ № 14-21-00158 «Разработка алгоритмов и методов для задач математического моделирования на суперкомпьютерных системах, включая гибридные» (руководитель — Шур Л. Н.)

На конференции, материалы которой публикуются в настоящем сборнике, представлено 23 устных и стендовых доклада.

Сотрудник Института космических исследований РАН (Москва) Моисеенко Сергей Григорьевич представил доклад «Изоэнтропические „ударные волны“ при численном моделировании в астрофизических задачах» [10], в котором развит численный метод, основанный на сохранении энтропии, позволяющий получить результаты в тех случаях, когда методы, основанный на сохранении энергии, неустойчив. Такие результаты, однако, имеют небольшую точность и могут быть использованы только для качественного понимания моделируемых астрофизических процессов.

Доклад Рыбакина Виктора Петровича [11], сотрудника МГУ им. М. В. Ломоносова (Москва), был посвящён моделированию процессов фрагментации молекулярного облака ударной волной. Проведённые расчёты позволили обосновать образование вихрей и исследовать абляцию вещества молекулярного облака.

Дружинин Олег Александрович (Институт прикладной физики РАН, Нижний Новгород) привёл результаты прямого моделирования турбулентного пограничного слоя над поверхностью океана, в которой обнаружено усиление скорости воздушного потока за счёт отрыва водных капель от поверхности волн [12].

Аспирантка кафедры теоретической физики физического факультета Нижегородского государственного университета Наталья Викторовна Дербенева [13] представила доклад о моделировании легирования нанокластеров кремния. Эта работа имеет перспективы в связи с поиском новых материалов для электроники.

Доклад аспиранта кафедры теоретической физики физического факультета Омского государственного университета Ивана Сергеевича Попова был основан на подготовленной им кандидатской диссертации, посвящённой численному исследованию методом Монте Карло структурно неупорядоченных спиновых моделей. Исследования были проведены с использованием российских суперкомпьютеров, таких как «Ломоносов» (МГУ) и «Чебышев» (ННГУ) [14].

Традиционно в первый день конференции проводится доклад во время ужина. Время доклада формально не ограничено. Доклад, как правило, имеет обзорный характер по интересной и важной тематике. Была выбрана тема, связанная с квантовыми вычислениями. Заведующий кафедрой теоретической физики физического факультета ННГУ профессор Сатанин Аркадий Михайлович начал доклад с исторического анализа развития вычислительной техники, от абака до аналогового квантового компьютера D-Wave. Слушатели получили возможность ознакомиться с основными идеями квантовых компьютеров, как основанных на квантовых операциях, так и на аналоговых принципах [15].

Доклад Глызина Сергея Дмитриевича был посвящен изложению модели для автоволновых процессов в кольцевых цепочках [16]. На основе этого доклада написана статья сборника [17]. В сборник включена статья коллектива, которая была доложена на предыдущем научном мероприятии в Тарусе в апреле 2015 г., но не вошла во вторую часть сборника по техническим причинам [31].

В работе конференции принял участие гость Научного центра РАН в Черноголовке из Словакии Михал Боровский. Он продемонстрировал успешность метода Parallel Annealing для нахождения основного состояния фрустрированной антиферромагнитной модели Изинга на стопке треугольных решёток, в то время как традиционные методы Монте Карло не позволяют получить основное состояние [18].

Лев Юрьевич Бараш из Черноголовки поделился опытом использования технологии OpenACC для написания программ с использованием ускорителей вычислений. Технология годится для работы с графическими ускорителями всех производителей, а также для ускорителя технологии Xeon Phi фирмы Инテル [19].

Аспирант Фроленко Иван Олегович (Ярославский государственный университет им. П.Г. Демидова) представил доклад на техническую тему — «Декомпозиция вычислительных сеток для гетерогенных кластерных архитектур» [20].

Молодой сотрудник НИВЦ МГУ им. М.В. Ломоносова Сергей Жуматий сделал интересный доклад об системном программном обеспечении, разрабатываемом для оптимизации загрузки суперкомпьютеров «Ломоносов» и «Ломоносов-2» [21].

Марат Галимов представил обзор проводимых коллективом исследователей из Казанского федерального университета работ по разработке суперкомпьютерного комплекса для решения обратных задач сейсморазведки [22].

Проблемам решения транспортной проблемы города Москвы был посвящён доклад Алексея Эдуардовича Воробьева, сотрудника Института Генплана Москвы. Участники конференции с интересом познакомились с фактическими данными о транспортных потоках и теми моделями, с помощью которых планируется развитие транспортной системы столицы [23].

В серии докладов аспирантов и студентов Ярославского государственного университета обсуждались методы и подходы исследования нелинейных задач

с запаздыванием [24–26]. Оригинальные подходы успешно разрабатываются коллективом в последние годы и с интересом принимаются участниками семинаров и конференций.

Последний день конференции был посвящён проблемам разработки программного обеспечения для использования в научных исследованиях. Вениамин Борисович Коношлев изложил подход коллектива Института космических исследований к хранению и обработке экспериментальных данных большого объёма [27]. Александр Павлович Крюков поделился опытом коллектива МГУ по разработке веб-платформы [28]. Василий Николаевич Храмушин подробно изложил свой подход к разработке программного обеспечения для исследования задач гидродинамики судов [29].

Работа конференции завершилась докладом Льва Николаевича Щура, в котором он проанализировал проблему сохранения ИКТ инфраструктуры учреждений РАН и её использовании для создания распределённой программно-аппаратной платформы для работы с данными большого объёма [30]. Участники конференции поддержали предлагаемые в докладе методы решения проблемы.

С полными мультимедийными версиями докладов можно ознакомиться на странице семинара [32].

Литература

- [1] Труды семинара по вычислительным технологиям в естественных науках. Вып. 1. Вычислительная физика / Под ред. Р. Р. Назирова. М.: КДУ, 2009.
- [2] Методы нелинейного анализа в кардиологии и онкологии: Физические подходы и клиническая практика. Вып. 2 / Под ред. Р. Р. Назирова. М.: КДУ, 2010.
- [3] Труды семинара по вычислительным технологиям в естественных науках. Вып. 3. Вычислительная химия и биология / Под ред. Р. Р. Назирова, Л. Н. Щура. М.: КДУ, 2010.
- [4] Техническое зрение в системах управления мобильными объектами-2010: Труды научно-технической конференции-семинара. Вып. 4. / Под ред. Р. Р. Назирова. М.: КДУ, 2011. 328 с.
- [5] Вычислительные технологии в естественных науках. Системы глобального масштаба: Труды семинара / Под ред. Р. Р. Назирова, Л. Н. Щура. М.: КДУ, 2011.
- [6] Вычислительные технологии в естественных науках. Перспективные компьютерные системы: устройства, методы и концепции: Труды семинара / Под ред. Р. Р. Назирова, Л. Н. Щура. М.: ИКИ РАН, 2011.
- [7] Инфокоммуникационные технологии в научных исследованиях // Информационное общество. 2013. № 1-2.
- [8] Вычислительные технологии в естественных науках. Методы суперкомпьютерного моделирования: Труды семинара / Под ред. Р. Р. Назирова, Л. Н. Щура. М.: ИКИ РАН, 2014.
- [9] Вычислительные технологии в естественных науках. Методы суперкомпьютерного моделирования. Часть 2. Труды семинара / Под ред. Р. Р. Назирова, Л. Н. Щура. М.: ИКИ РАН, 2015.
- [10] *Моисеенко С. Г.* Изоэнтропические «ударные волны» при численном моделировании в астрофизических задачах: доклад.
- [11] *Рыбакин Б. П.* Математическое моделирование процессов фрагментации и абляции молекулярного облака сильной ударной волной (см. статью настоящего сборника, с. 161).
- [12] *Дружинин О. А., Троицкая Ю. И.* Прямое численное моделирование двухфазного турбулентного пограничного слоя над взволнованной водной поверхностью (статья настоящего сборника, с. 114).
- [13] *Дербенева Н. В.* Влияние фосфора на Оже-рекомбинацию в нанокристаллах кремния (статья настоящего сборника, с. 107).

- [14] *Попов И. С., Прудников П. В., Прудников В. В.* Численное описание неравновесного критического поведения двумерной структурно неупорядоченной XY-модели методами Монте-Карло (статья настоящего сборника, с. 150).
- [15] *Сатанин А. М.* Адиабатические вычисления, «квантовый отжиг» и неравновесная термодинамика: доклад.
- [16] *Глызин С. Д.* Автоволновые процессы в кольцевых цепях с однонаправленной связью: доклад.
- [17] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Режимы двухкластерной синхронизации в полностью связанной сети релаксационных осцилляторов: доклад.
- [18] *Боровский М.* Population annealing study of the frustrated Ising antiferromagnet on the stacked triangular lattice: доклад.
- [19] *Бараш Л. Ю.* Оптимизация алгоритмов численного решения задач магнитогидродинамики в ОТО для GPU при помощи технологии OpenACC: доклад.
- [20] *Фроленко И. О.* Декомпозиция вычислительных стоек для гетерогенных кластерных архитектур: доклад.
- [21] *Леоненков С. Н., Жуматий С.* Расширение алгоритма Backfill для использования на суперкомпьютере «Ломоносов» (статья настоящего сборника, с. 140).
- [22] *Биряльцев Е. В., Галимов М. Р., Демидов Д. Е., Елизаров А. М.* Опыт разработки и эксплуатации суперкомпьютерного комплекса для решения обратных задач сейсморазведки (статья настоящего сборника, с. 18).
- [23] *Воробьев А. Э., Титов А. Ю., Гаврилин В. А., Меньшиутин А. Ю., Бахирев И. А.* Транспортная модель Московского региона (статья настоящего сборника, с. 49).
- [24] *Алешин С. В.* Качественная структура процесса распространения возмущений в уравнении Колмогорова-Петровского-Пискунова с отклонением по пространственной переменной (статья настоящего сборника, с. 9).
- [25] *Горюнов В. Е.* Пространственно неоднородные режимы логистического уравнения с запаздыванием и малой диффузией в плоской области: доклад.
- [26] *Ивановский Л. И.* Динамические свойства одного класса импульсных систем (статья настоящего сборника, с. 126).
- [27] *Коноплев В. В., Боярский М. Н., Назиров Р. Р.* SCARP: масштабируемая и отказоустойчивая платформа лёгкой виртуализации для интенсивной работы с большими данными и облачных приложений (статья настоящего сборника, с. 132).
- [28] *Крюков А. П., Демичев А.* Разработка веб-платформ для предоставления прикладного ПО как веб-сервисов (SaaS) в подходе открытого рынка: доклад.
- [29] *Богданов А. В., Дегтярев А. Б., Храмушин В. Н.* Трёхмерная тензорная математика вычислительных экспериментов в гидромеханике (статья настоящего сборника, с. 34).
- [30] *Шур Л. Н.* К вопросу об информационно-телекоммуникационной инфраструктуре, а также о возможности проведения учреждениями ФАНО исследований в области Интенсивной обработки данных с целью извлечения нового знания (статья настоящего сборника, с. 186).
- [31] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Неклассические релаксационные колебания и бифуркации голубого неба в задаче моделирования кардиоритмов (статья настоящего сборника, с. 63).
- [32] Мультимедийные версии докладов. URL: <http://www.iki.rssi.ru/seminar/2015111719/>.

Р. Р. Назиров,
доктор технических наук, ИКИ РАН
Л. Н. Шур
профессор, доктор физико-математических наук,
ИТФ РАН, Научный центр РАН в Черноголовке

КАЧЕСТВЕННАЯ СТРУКТУРА ПРОЦЕССА РАСПРОСТРАНЕНИЯ ВОЗМУЩЕНИЙ В УРАВНЕНИИ КОЛМОГОРОВА-ПЕТРОВСКОГО-ПИСКУНОВА С ОТКЛОНЕНИЕМ ПО ПРОСТРАНСТВЕННОЙ ПЕРЕМЕННОЙ

С. В. Алешин

Ярославский государственный университет им. П. Г. Демидова (ЯрГУ им. П. Г. Демидова)
Научный центр Российской академии наук в Черноголовке

В работе рассматривается задача распространения волны плотности в логистическом уравнении с отклонением пространственного аргумента и диффузией (уравнение Фишера-Колмогорова-Петровского-Пискунова с отклонением). Проведён численный анализ процесса распространения волны. Выделены значения отклонения, для которых характерны качественно различные пространственно-временные распределения решений.

Исследование выполнено поддержке Российского научного фонда (проект № 14-21-00158).

Ключевые слова: уравнение Фишера-Колмогорова-Петровского-Пискунова, диффузия, отклонение, уравнения с запаздыванием, нормальные формы, численные методы.

ВВЕДЕНИЕ

Одним из наиболее простых нелинейных уравнений в частных производных, моделирующих распространение волн плотности распределения изучаемой величины, является рассмотренное А. Н. Колмогоровым, И. Г. Петровским и Н. С. Пискуновым в статье [Колмогоров и др., 1937] логистическое уравнение с диффузией

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + u(1-u), \quad (1)$$

которое ниже будем называть уравнением КПП. Это уравнение было использовано авторами для описания процесса распространения генной волны. Здесь $u(t, x)$ — плотность распределения числа особей, обладающих доминантным геном; $t \geq 0$ — временная переменная; $x \in (-\infty, +\infty)$ — пространственная. Рассматривается задача распространения волны плотности численности от некоторого ненулевого начального условия. Почти одновременно с работой [Колмогоров и др., 1937] появилась статья Р. А. Фишера [Fisher, 1937], также посвящённая анализу волн плотности численности в уравнении (1), в связи с чем его часто называют уравнением Фишера или Фишера-Колмогорова-Петровского-Пискунова.

Уравнение (1) находит применение в широкой области приложений, связанных с распространением в пространстве волн различной природы от концентрации некоторого реагирующего вещества до волн плотности популяции. В работах [Колмогоров и др., 1937; Fisher, 1937] рассмотрен вопрос о таких неотрицательных решениях $u(t, x)$, для которых при каждом $t > 0$ выполнены условия:

$u(t, x) \rightarrow 1$ при $x \rightarrow +\infty$ и $u(t, x) \rightarrow 0$ при $x \rightarrow -\infty$. Показано, что при достаточно больших t функция $u(t, x)$ принимает некоторую фиксированную форму.

Обобщения уравнения (1) могут быть связаны с учётом зависимостей нелинейных слагаемых этого уравнения от сдвигов по пространству и по времени. В работах [Колесов, Колесов, 1993] логистическое уравнение (1) с различными краевыми условиями и запаздывающим аргументом было применено к моделированию динамики плотности популяции. Несколько более общий подход продемонстрирован в статье [Гурли и др., 2003], где дана развёрнутая постановка задачи, в которой изучается пространственное распределение плотности популяции. В частности, предлагается использовать следующее уравнение (см. также работу Н. Бриттона [Britton, 1990]):

$$\frac{\partial N(t, x)}{\partial t} = \Delta N(t, x) + N(t, x)(1 + \alpha N(t, x) - (1 - \alpha)(g * N)(t, x)), \quad (2)$$

где $N(t, x)$ — плотность популяции в момент времени t в точке x ареала обитания; выражение $1 + \alpha N(t, x) - (1 - \alpha)(g * N)(t, x)$ определяет изменения плотности численности; Δ — оператор Лапласа; функция g характеризует пространственно-временные неоднородности. В статье [Britton, 1990] предлагается следующий вид свёртки $(g * N)$

$$(g * N)(t, x) = \int_{-\infty}^t \int_{\Omega} g(t - \tau, x - y) N(\tau, y) dy d\tau, \quad (3)$$

где Ω — область распределения популяции. Функция g должна удовлетворять условию нормировки $(g * 1)(t, x) = 1$. В предельном случае, когда g представляет собой дельта-функцию, сосредоточенную в начале координат, получаем уравнение Фишера-Колмогорова-Петровского-Пискунова (1). Если точку сосредоточения сдвинуть по оси времени, то получится уравнение с запаздыванием, подробно рассмотренное в статье авторов [Алешин и др., 2015], если же точку сосредоточения сдвинуть по пространственной переменной, то получится уравнение с отклонением по пространству.

Рассмотрим обобщение уравнения КПП (1), содержащее отклонение по пространству

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + u(1 - u(t, x - h)). \quad (4)$$

Величину h пространственного отклонения будем считать положительным числом, поскольку в противном случае можно поменять переменную x на $-x$, что приводит к смене знака параметра отклонения.

Исследование качественного поведения решений уравнения (4) открывается в первом разделе статьи анализом уравнения профиля волны, для которого найдены условия возникновения колебательных режимов.

Затем, во втором разделе, проанализирована соответствующая уравнению (4) краевая задача с периодическими условиями, для которой изучена проблема потери устойчивости пространственно однородного состояния равновесия и найдены ответвляющиеся от него пространственно неоднородные колебательные режимы.

Учитывая, что для многих приложений представляет интерес задача о распространении волн концентрации в задаче (4), в последней части работы приведены результаты численного моделирования для уравнения (4) в случае неограниченной по x области.

НЕКОТОРЫЕ СВОЙСТВА ВОЛНОВЫХ РЕШЕНИЙ ЗАДАЧИ

В работе [Колмогоров и др., 1937] было показано, что в уравнении (1) волны распространяются вдоль направлений $2t \pm x = \text{const}$ и был определён профиль волны, приводящей к переходу от нулевых значений переменной к единичным. Как и в работе [Алешин и др., 2015] выполним в уравнении (4) замену в виде бегущей волны вида $u(t, x) = w(2t \pm x)$ и перейдём к новой временной переменной $s = 2t \pm x$, тогда для переменной $w(s)$ имеем следующее уравнение второго порядка с запаздыванием:

$$w'' - 2w' + w(1 - w(s - h)) = 0, \quad (5)$$

где штрихом обозначена производная по переменной s . Свойства устойчивости нулевого решения уравнения (5) не зависят от h , это решение представляет собой неустойчивый узел с кратным корнем, равным единице. Свойства устойчивости единичного состояния равновесия определяются расположением корней характеристического квазиполинома

$$P(\lambda) \equiv \lambda^2 - 2\lambda - \exp(-h\lambda). \quad (6)$$

Доказаны следующие утверждения.

Лемма 1. *Квазиполином $P(\lambda)$ имеет при $0 < h < h_1^*$ ровно три вещественных корня: один положительный и два отрицательных, а при $h > h_1^*$ — единственный положительный вещественный корень. Здесь $h_1^* = -(2\lambda_* - 2)/(\lambda_*^2 - 2\lambda_*) \approx 1,12154$; λ_* является корнем трансцендентного уравнения $\lambda^2 - 2\lambda - \exp((2 - 2\lambda)/(\lambda - 2)) = 0$.*

Лемма 2. *Все корни квазиполинома $P(\lambda)$ кроме одного вещественного положительного лежат при $0 < h < h_2^*$ в левой комплексной полуплоскости. Здесь*

$$h_2^* = \frac{\arccos\left(\frac{-\sqrt{5} + 2}{\sqrt{5} - 2}\right)}{\sqrt{5} - 2} \approx 3,72346. \quad (7)$$

При $h = h_2^*$ на мнимую ось выходит пара чисто мнимых корней $\lambda = \pm i\omega_0$, причём

$$\omega_0 = \sqrt{\sqrt{5} - 2} \approx 0,48587. \quad (8)$$

В окрестности решения $w(s) = 1$ найдена асимптотика цикла, ответвляющегося от этого решения при $h = h_2^* + \mu$, где $0 < \mu \ll 1$. В этом случае верно следующее утверждение.

Лемма 3. *Пусть $h = h_2^* + \mu$, где $0 < \mu \ll 1$, тогда существует такое $\mu_0 > 0$, что для всех $0 < \mu < \mu_0$ уравнение (5) имеет дихотомичный цикл, неустойчивое многообразие которого одномерно, а асимптотика задаётся формулой*

$$w(s, \mu) = 1 + \sqrt{\mu} \left(z(\tau) \exp(i\omega_0 s) + \bar{z}(\tau) \exp(-i\omega_0 s) \right) + \mu w_1(s, \tau) + \mu^{3/2} w_2(s, \tau) + \dots,$$

$\tau = \mu s$, $w_j(s, \tau)$ ($j = 1, 2$), в которой комплексная амплитуда $z(\tau)$ заменена выражением

$$\sqrt{\frac{\text{Re}(\varphi_0)}{\text{Re}(\varphi_1)}} \exp \left| \frac{i\epsilon s (\text{Im}(\varphi_0) \text{Re}(\varphi_1) - \text{Re}(\varphi_0) \text{Im}(\varphi_1))}{\text{Re}(\varphi_0)} + i\gamma \right|,$$

$$\text{и } \varphi_0 = \frac{2\omega_0^2(-1 + i\omega_0)}{P'(i\omega_0)}; \quad \varphi_1 = \frac{1}{P'(i\omega_0)} \left[2\omega_0^2(1 - \omega_0^2 - 2i\omega_0) + \beta \left| \frac{\omega_0^2 + 2i\omega_0}{-1/(\omega_0^2 + 2i\omega_0)} \right| \right];$$

$$\beta = \frac{\omega_0^2 + 2i\omega_0}{4\omega_0^2 + 4i\omega_0 + (\omega_0^2 + 2i\omega_0)^2}; \quad \gamma - \text{произвольная константа, определяющая фазовый}$$

сдвиг вдоль цикла. Приближенные значения коэффициентов следующие: $\Phi_0 \approx 0,136807 - 0,20660i$, $\Phi_1 \approx -0,04429 - 0,03664i$.

ВОЛНОВЫЕ РЕШЕНИЯ В ЗАДАЧЕ С ПЕРИОДИЧЕСКИМИ УСЛОВИЯМИ

В этом разделе рассмотрим логистическое уравнение с диффузией и отклонением (4), дополненное периодическими краевыми условиями

$$u(t, x) = u(t, x + T), \quad (9)$$

где $T > 0$ — период. В этой ситуации фазовым пространством задачи (4), (9) будем считать соболевское пространство T -периодических функций $W_2^2(0, T)$. Для исследования устойчивости состояния равновесия $u(t, x) = 1$ краевой задачи (4), (9) линеаризуем её на этом решении, приходим к краевой задаче (аналогичные действия для задачи с запаздыванием выполнялись в работах [Глызин, 2009, 2010; Глызин и др., 2010; Кашенко, 1987, 1990; Glyzin, 2013; Kashchenko, 2013; Wu, 1996]).

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} - v(t, x - h), \quad v(t, x) = v(t, x + T). \quad (10)$$

Если в выражениях (10) выполнить разложение по пространственным модам, то на каждой из мод после замены $v(t, x) = \exp \lambda \exp i\omega x$ приходим к уравнению

$$\lambda = -\omega^2 - \exp i\omega h. \quad (11)$$

Нетрудно видеть, что при достаточно малых h величина $\operatorname{Re} \lambda < 0$. Найдём первое положительное значение параметра h , при котором $\operatorname{Re} \lambda = 0$. Для этого необходимо решить алгебраическую систему

$$-\omega^2 - \cos \omega h = 0, \quad -2\omega + h \sin \omega h = 0. \quad (12)$$

Эта система сводится к трансцендентному уравнению

$$-\omega h = 2 \operatorname{ctg}(\omega h), \quad (13)$$

решая которое относительно ωh , получаем решение системы (15):

$$h^* = 2,791544, \quad \omega^* = 0,88077. \quad (14)$$

Нетрудно заметить, что при $h < h^*$ состояние равновесия $u(t, x) = 1$ краевой задачи (4), (9) асимптотически устойчиво. Предположим теперь, что $T = 2\pi/\omega^*$, а $h = h^* + \varepsilon$, где ε положительно и мало, и выясним характер потери устойчивости этого состояния равновесия. Для этого выполним в (4), (9) стандартную замену метода нормальных форм

$$u(t, x, \varepsilon) = 1 + \sqrt{\varepsilon} \cdot u_0(t, \tau, x) + \varepsilon u_1(t, \tau, x) + \varepsilon^{3/2} u_2(t, \tau, x) + \dots, \quad (15)$$

где $\tau = \varepsilon t$ — медленное время;

$$\omega_0 = \sin \omega^* h^*; \quad u_0(t, \tau, x) = z(\tau) \exp[i(\omega_0 t + \omega^* x)] + \bar{z}(\tau) \exp[-i(\omega_0 t + \omega^* x)].$$

Приравнявая коэффициенты при одинаковых степенях $\sqrt{\varepsilon}$ и отыскивая $u_j(t, \tau, x)$, $j = 1, 2$ в виде тригонометрических многочленов по t , на втором шаге получаем функцию

$$u_1(t, x, \varepsilon) = -2|z|^2 \cos \omega^* h^* + \left\{ z^2 w_2 \exp[2i(\omega_0 t + \omega^* x)] + \text{к.с.} \right\}, \quad (16)$$

где под к.с. подразумевается выражение, комплексно сопряжённое с данным в той же скобке; $w_2 = -\left(2\omega_0 i + 4\omega^* + \exp(-2i\omega^* h^*)\right)^{-1} \exp(-i\omega^* h^*)$.

На третьем шаге выполнения алгоритма в краевой задаче для $u_2(t, \tau, x)$ из условий её разрешимости в классе тригонометрических многочленов по t получаем следующее укороченное уравнение нормальной формы:

$$\frac{dz}{d\tau} = \varphi_0 z + \varphi_1 |z|^2 z, \quad (17)$$

где $\varphi_0 = i\omega^* \exp(-i\omega^* h^*)$;

$$\varphi_1 = 2\cos\omega^* h^* (1 + \exp(-i\omega^* h^*)) - (\exp(-2i\omega^* h^*) + \exp(i\omega^* h^*)) w_2.$$

Используя полученное представление и формулу (14), нетрудно найти приближенные значения коэффициентов нормальной формы: $\varphi_0 \approx 0,5558 - 0,6833i$, $\varphi_1 \approx -0,1701 + 0,59i$.

Полученные результаты позволяют сформулировать следующее утверждение.

Лемма 4. Пусть $h = h^* + \varepsilon$, тогда существует такое $\varepsilon_0 > 0$, что для всех $0 < \varepsilon < \varepsilon_0$ краевая задача (4), (9) имеет орбитально асимптотически устойчивый цикл, асимптотика которого задаётся формулой (15), где медленная переменная z заменена выражением

$$z = \sqrt{\frac{\operatorname{Re}(\varphi_0)}{\operatorname{Re}(\varphi_1)}} \exp\left\{ \frac{i\varepsilon t (\operatorname{Im}(\varphi_0) \operatorname{Re}(\varphi_1) - \operatorname{Re}(\varphi_0) \operatorname{Im}(\varphi_1))}{\operatorname{Re}(\varphi_0)} + i\gamma \right\} \quad (18)$$

(γ — произвольная константа, определяющая фазовый сдвиг вдоль цикла).

ЧИСЛЕННЫЙ АНАЛИЗ УРАВНЕНИЯ КОЛМОГОРОВА-ПЕТРОВСКОГО-ПИСКУНОВА С ПРОСТРАНСТВЕННЫМ ОТКЛОНЕНИЕМ

Численное исследование распространения волн концентрации в уравнении (4) от локализованного по пространству начального возмущения выполнялось на некотором отрезке $[a, b]$. При этом разница $|a - b|$ выбиралась достаточно большой для того, чтобы можно было проследить за распространением волны от начального момента до момента встречи фронта волны с границами a или b . Учитывая, что рассматривается задача о распространении локального возмущения, зададим на границах отрезка нулевые краевые условия $u(t, a) = u(t, b) = 0$.

Перейдём к описанию вычислительной процедуры. Вторую производную по пространственной переменной в правой части уравнения (4) заменим конечной разностью второго порядка. Для этого отрезок $[a, b]$ разобьём на N равных частей и построим сетку узлов с шагом $\Delta x = (b - a)/N$ так, что $x_j = a + j\Delta x$, где $j = 0, \dots, N - 1$. Обозначим через $u_j(t)$ значение функции $u(t, x)$ в соответствующих узлах сетки. В итоге получим следующую систему обыкновенных дифференциальных уравнений:

$$\dot{u}_j = \frac{u_{j+1} - 2u_j + u_{j-1}}{(\Delta x)^2} + (1 - u_{j-k})u_j, \quad j = 0, \dots, N - 1, \quad (19)$$

где $k = \lfloor h/\Delta x \rfloor$ отвечает за пространственное отклонение (здесь $\lfloor \cdot \rfloor$ обозначена целая часть числа). Для учёта краевых условий полагаем $u_{-1}(t) = u_N(t) = 0$.

Численный эксперимент выполнялся на вычислительном кластере ЯрГУ (Международной научно-исследовательской лаборатории «Дискретная и вычис-

лительная геометрия» им. Б. Н. Делоне). Одновременно решалось от $N = 1,8 \cdot 10^5$ до $N = 1,8 \cdot 10^6$ обыкновенных дифференциальных уравнений. Для вычислений использовался метод Дормана-Принца пятого порядка с переменной длиной шага интегрирования. Абсолютная и относительная погрешности алгоритма были приняты равными 10^{-12} . Начальный шаг интегрирования взят равным 10^{-3} .

Начальные условия выбирались в виде прямоугольного импульса высоты 0,1 и единичной ширины, расположенного в центре отрезка $[a, b]$. В частности, для случая $x \in [0, 1800]$:

$$u_j(0) = \begin{cases} 0,1, & \text{если } j \in [89950, 90050] \\ 0, & \text{если иначе.} \end{cases} \quad (20)$$

Для последующего анализа и графического отображения полученные данные прореживались.

Перейдём к результатам численного моделирования. Описание поведения решений уравнения (4) с отклонением пространственной переменной будем проводить в сравнении с классическим уравнением КПП без отклонения. На рис. 1 представлено распространение волны постоянной высоты от начального всплеска единичной ширины и высоты 0,1 при $h = 1,2$. Скорость распространения волны согласно статье [Колмогоров и др., 1937] равнялась двум (на соответствующих рисунках скорость распространения равна углу наклона профиля волны). Нетрудно заметить, что при этих значениях отклонения график процесса распространения волны для задачи с отклонением и без отклонения практически не отличается.

При увеличении параметра h можно выделить несколько этапов качественно различного поведения решений системы (19).

1. При относительно малом значении h на промежутке от нуля до h_1^* поведение системы (19) практически неотличимо от поведения системы КПП без отклонения. Величина h_1^* ограничивает данный промежуток, по-видимому, в связи с тем, что решения уравнения (5) на устойчивом инвариантном многообразии единичного состояния равновесия монотонно стремятся к этому состоянию (см. рис. 1 при $h = 1,2$).

2. При $h_1^* < h < h^*$ фронт распространения волны приближается к единичному значению колебательным образом. На рис. 2 показана такая волна при $h = 2,7$. В этом случае максимальная амплитуда всплеска равна примерно 1,5, а после прохождения фронта волны значение $u(t, x)$ быстро приближается к единице.

3. Следующее существенное изменение в распространении фронта волны уравнения (4) наблюдается при $h > h^*$. Существенной особенностью решения в этом случае является то, что оно перестаёт удовлетворять уравнению (5). Это происходит в силу того, что решение $u = 1$ периодической красной задачи (4), (9) колебательно теряет устойчивость и от него отщепляется орбитально асимптотически устойчивый, пространственно неоднородный цикл. Указанное обстоятельство приводит к тому, что в правой части области распространения волны появляется расширяющийся с течением времени участок с интенсивными пространственными колебаниями. На рис. 3 показано распространение волны при $h = 2,81$, при этом на рис. 4 приведена зависимость $u(t, x)$ при $t = 4500$ и изменении пространственной переменной x в пределах от 0 до 18 000. Структура колебательного процесса, представленного на данных графиках, видимо, является установившейся. Во всяком случае, к моменту времени $t = 4500$ амплитуда колебаний не растёт с ростом t и в целом сохраняется довольно сложная пространственная структура решения $u(t, x)$.

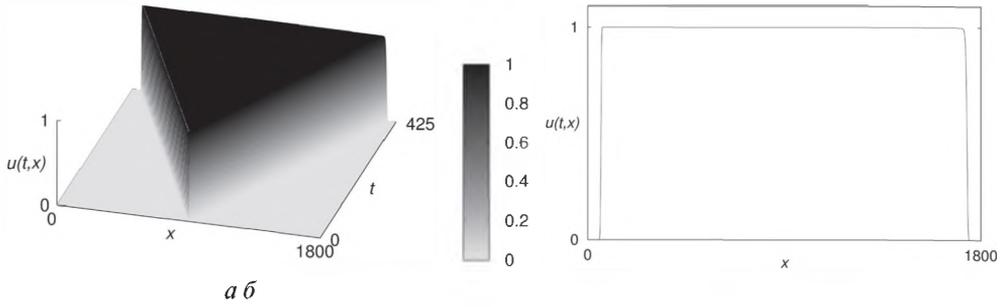


Рис. 1. Решение системы (19) с начальными условиями (20) при отклонении $h = 1,2$:
 a — распространение волны на плоскости (x, t) ; b — разрез при $t = 420$

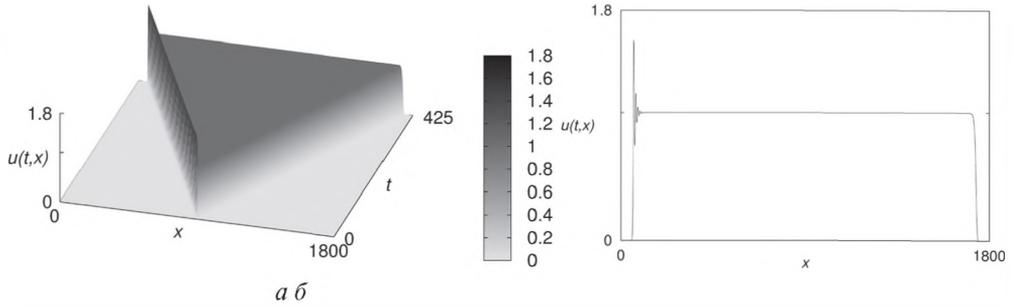


Рис. 2. Решение системы (19) с начальными условиями (20) при отклонении $h = 2,7$:
 a — распространение волны на плоскости (x, t) ; b — разрез при $t = 420$

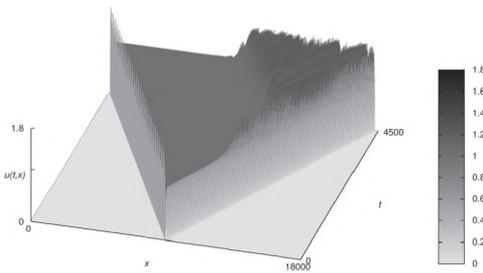


Рис. 3. Распространение волны в системе (19) при $h = 2,81$

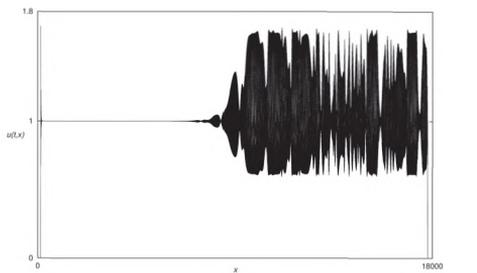


Рис. 4. Пространственное распределение решения системы (19) при $t = 4500$ и отклонении $h = 2,81$

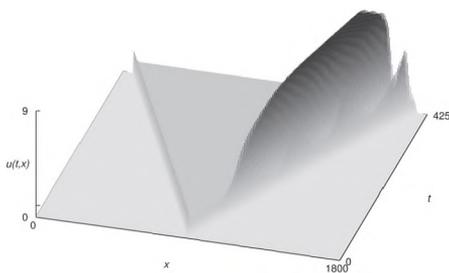


Рис. 5. Распространение волны в системе (19) при $h = 3$

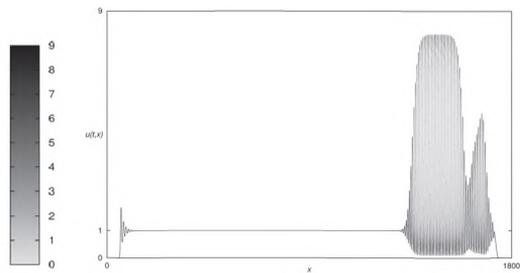


Рис. 6. Пространственное распределение решения системы (19) при $t = 425$ и отклонении $h = 3$

4. Численный эксперимент не удалось продолжить для значений отклонения h , существенно отличающихся от h^* . Так, уже при $h = 3$ в правой части графика распространения волны возникают весьма интенсивные колебания, достигающие при $t = 425$ значения около 9. На рис. 5 показано распространение волны при $h = 3$, при этом на рис. 6 приведена зависимость $u(t, x)$ при $t = 425$ и $x \in [0, 1800]$.

Дальнейшее изменение отклонения приводит к резкому увеличению амплитуды колебательного режима, что с одной стороны, влечёт необходимость дробления шага по времени для сохранения точности вычислений, а с другой — значительно замедляет вычислительный процесс. По указанным причинам вычислительный эксперимент не был распространён на большие значения отклонения.

ЗАКЛЮЧЕНИЕ

Распространение возмущения в активной среде при учёте отклонения по пространственной переменной представляет собой сложный процесс. Эта сложность обуславливает необходимость применения различных сочетаний аналитических и численных методов. На этом пути были получены следующие результаты.

Изучение уравнения распространения волны (5) позволило найти критические значения параметра запаздывания, при которых меняется вид пространственного распределения решения задачи.

Для выяснения некоторых особенностей качественного поведения решений уравнения КПП с отклонением было изучено уравнение (4) с периодическими краевыми условиями вблизи состояния равновесия $u_0 = 1$. Это позволило выяснить характер потери устойчивости пространственно однородного решения этой задачи и найти асимптотику режимов, возникающих при этом. Численный анализ, выполненный с учётом аналитических результатов, позволил выделить следующие промежутки значений отклонения:

- на которых поведение решений уравнения с отклонением близко к их поведению в задаче без отклонения;
- при которых на левом фронте области распространения волны появляются участки с колебаниями, затухающими к единице;
- при которых вся область распространения волны заполняется интенсивными колебаниями решения со сложным пространственным распределением.

Автор выражает благодарность Глызину С.Д и Кащенко С.А. за постановку задачи и обсуждение полученных результатов.

ЛИТЕРАТУРА

- [Алешин и др., 2015] Алешин С. В., Глызин С. Д., Кащенко С. А. Уравнение Колмогорова-Петровского-Пискунова с запаздыванием // Моделирование и анализ информационных систем. 2015. Т. 22. № 2. С. 304–321.
- [Глызин, 2009] Глызин С. Д. Разностные аппроксимации уравнения «реакция-диффузия» на отрезке // Моделирование и анализ информационных систем. 2009. Т. 16. № 3. С. 96–116.
- [Глызин и др., 2010] Глызин С. Д., Колесов А. Ю., Розов Н. Х. Конечномерные модели диффузионного хаоса // Журн. вычислит. математики и мат. физики. 2010. Т. 50. № 5. С. 860–875.

- [Гурли и др., 2003] *Гурли С. А., Соу Дж. В.-Х., Ву Дж. Х.* Нелокальные уравнения реакции-диффузии с запаздыванием: биологические модели и нелинейная динамика // Современная математика. Фундаментальные направления. 2003. Т. 1. С. 84–120.
- [Кашенко, 1987] *Кашенко С. А.* Об установившихся режимах уравнения Хатчинсона с диффузией // ДАН СССР. 1987. Т. 292. № 2. С. 327–330.
- [Кашенко, 1990] *Кашенко С. А.* Пространственно-неоднородные структуры в простейших моделях с запаздыванием и диффузией // Мат. моделирование. 1990. Т. 2. № 9. С. 49–69.
- [Колесов, Колесов, 1993] *Колесов А. Ю., Колесов Ю. С.* Релаксационные колебания в математических моделях экологии // Тр. МИАН. Т. 199. М.: Наука, 1993. 126 с.
- [Колмогоров и др., 1937] *Колмогоров А. Н., Петровский И. Г., Пискунов Н. С.* Исследование уравнения диффузии, соединённой с возрастом вещества, и его применение к одной биологической проблеме // Бюл. МГУ. Сер. А. Математика и механика. 1937. Т. 1. № 6. С. 1–26.
- [Britton, 1990] *Britton N. F.* Spatial structures and periodic travelling waves in an integro-differential reaction-diffusion population model // SIAM J. Appl. Math. 1990. V. 50. P. 1663–1688.
- [Fisher, 1937] *Fisher R. A.* The Wave of Advantageous Genes // Annals of Eugenics. 1937. V. 7. P. 355–369.
- [Glyzin, 2013] *Glyzin S. D.* Dimensional Characteristics of Diffusion Chaos // Automatic Control and Computer Sciences. 2013. V. 47. No. 7. P. 452–469.
- [Kashchenko, 2013] *Kashchenko S. A.* Asymptotics of the Solutions of the Generalized Hutchinson Equation // Automatic Control and Computer Science. 2013. V. 47. No. 7. P. 470–494.
- [Wu, 1996] *Wu J.* Theory and Applications of Partial Functional Differential Equations. N. Y.: Springer-Verlag, 1996. 432 с. URL: <http://www.springer.com/gp/book/9780387947716>.

QUALITATIVE STRUCTURE OF PERTURBATIONS PROPAGATION PROCESS OF THE FISHER-KOLMOGOROV EQUATION WITH A DEVIATION OF SPATIAL VARIABLE

S. V. Aleshin

P. G. Demidov Yaroslavl State University (P. G. Demidov YSU)
Science Center in Chernogolovka, Russian Academy of Sciences (SCC RAS)

We considered the problem of density wave propagation of logistic equation with a deviation of spatial variable and diffusion (Fisher-Kolmogorov equation with a deviation of spatial variable). We performed a numerical analysis of wave propagation and identified the values of deviation which is characterized by the appearance of qualitative different spatio-temporal distribution of solutions.

Keywords: Fisher-Kolmogorov equation, diffusion, spatial deviation, delay differential equation, normal form, numerical analysis.

ОПЫТ РАЗРАБОТКИ И ЭКСПЛУАТАЦИИ СУПЕРКОМПЬЮТЕРНОГО КОМПЛЕКСА ДЛЯ РЕШЕНИЯ ОБРАТНЫХ ЗАДАЧ СЕЙСМОРАЗВЕДКИ

Е. В. Биряльцев¹, М. Р. Галимов¹, Д. Е. Демидов², А. М. Елизаров^{3,2}

¹ ООО «Градиент технолоджи», Казань

² Казанское отделение Межведомственного суперкомпьютерного центра Российской академии наук — Филиала Федерального научного центра «НИИ системных исследований РАН» (МСЦ РАН — филиал ФНЦ НИИСИ РАН)

³ Институт математики и механики им. Н. И. Лобачевского Казанского (Приволжского) федерального университета (ИММ КФУ)

В работе описан опыт создания комплексного программно-аппаратного решения для организации высокопроизводительных вычислений при решении современных геофизических задач на основе методов полноволновой инверсии. Обсуждаются проблемы создания массовых высокопроизводительных программных комплексов для широкого использования в промышленности. Приведён обзор существующих программно-технических решений.

Работа выполнена при финансовой поддержке РФФИ (проекты № 15-07-05380, 15-47-02343).

Ключевые слова: графические вычисления общего назначения, численное моделирование, распределённые вычисления, полноволновая инверсия, кластер GPU, геофизические исследования, платформа высокопроизводительных вычислений.

ВВЕДЕНИЕ

В настоящий момент развитие суперкомпьютерных вычислений достигло очередного качественного рубежа — производительность кластеров превысила планку в десятки петафлопсов, и научное сообщество обсуждает вопросы будущих экзафлопсных вычислений [Da Costa et al., 2015]. Повышение конкурентоспособности предприятий и экономики в целом во многом определяется активным массовым использованием суперкомпьютерных технологий [Бетелин и др., 2007]. В такой ситуации широкое внедрение суперкомпьютерных систем в промышленное производство становится ещё более актуальным, тем более, что такое внедрение наблюдается сегодня лишь в крупных корпорациях и стратегических отраслях.

В статье рассмотрены проблемы внедрения суперкомпьютерных технологий в традиционные сектора производства. В качестве примера представлен опыт разработки и эксплуатации суперкомпьютерного комплекса для геофизической компании.

Биряльцев Евгений Васильевич — генеральный директор, кандидат технических наук, igenbir@ya.ru

Галимов Марат Разифович — зам. генерального директора, кандидат технических наук, glmvmrt@gmail.com

Демидов Денис Евгеньевич — старший научный сотрудник, кандидат физико-математических наук, denis.demidov@gmail.com

Елизаров Александр Михайлович — заведующий кафедрой ИММ КФУ, доктор физико-математических наук, профессор, директор Казанского отделения МСЦ РАН — филиала ФНЦ НИИСИ РАН, amelizarov@gmail.com

1. ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ ВЫЧИСЛЕНИЯ В СЕЙСМОРАЗВЕДКЕ

Сегодня нефтегазовая промышленность характеризуется активным поиском и внедрением новых технологий, что объясняется, в частности, постоянным и значительным увеличением стоимости разведки и добычи углеводородного сырья. В этой ситуации происходит поиск новых методов геологоразведки, эффективных и одновременно мало затратных, которые позволили бы увеличить качество геологического прогноза и экономическую обоснованность принимаемых технических решений.

Среди перспективных методов моделирования в геологии можно выделить технологии полноволновой инверсии [Haffinger, 2013], которые позволяют восстанавливать свойства среды в любых геологических условиях, в отличие от технологий, основанных на базовом способе современной сейсморазведки ОГТ (общей глубинной точки), которые эффективны при плоскопараллельной структуре исследуемой геологической среды. Частным случаем полноволновой инверсии можно считать разработанную в России технологию пассивной низкочастотной сейсморазведки нефти и газа [Шабалин и др., 2013], при которой анализируются изменения характеристик естественных микросейсм над нефтегазовыми залежами. Эта технология основана на расчётах распространения сейсмических волн в сложных геологических средах на сеточных моделях с размерностями (в настоящее время) до 10^9 . Для промышленного её применения необходимо реализовывать расчёты численных моделей такой размерности со скоростью, сопоставимой с темпом выполнения производственных работ (десятки часов), и экономически оптимальным способом. Для решения этой задачи специалисты Казанской геофизической компании ЗАО «Градиент» совместно с Казанским филиалом Федерального научного центра «НИИ системных исследований РАН» (НИИСИ РАН) и научно-производственной компанией ООО «Градиент технолоджи» (Казань), начиная с 2005 г., провели исследования по поиску наиболее эффективных программно-аппаратных средств для выполнения массовых математических вычислений при моделировании в геологии. Рассматривались различные аппаратные и программные платформы. Было разработано специализированное программное обеспечение для расчётов, в том числе, с использованием кластеров GPU. Результатом работ стало создание двух программно-аппаратных комплексов на основе графических ускорителей компании AMD. Последний из них, недавно показавший по Linpack-тестам производительность в 35,6 Тфлоп/с, успешно используется геофизической компанией-заказчиком, а также занял 38-е место в рейтинге Top 50 суперкомпьютеров СНГ (28.09.2015). Кроме того, в рейтинге данный комплекс характеризуется максимальной производительностью на вычислительный узел, а наличие лишь 6 вычислительных узлов с 24 графическими картами позволяет считать его компактным и экономичным. Опыт разработки и внедрения комплекса показал, что созданная аппаратная база в настоящее время готова для промышленного внедрения. Непосредственный процесс сборки вычислительных мощностей, развёртывания и настройки базового программного обеспечения не представлял особой сложности. Кроме того, в настоящий момент на российском ИТ-рынке присутствуют производители (Г-Платформы, Meijin, IntellectDigital, ARBYTE SC, STSS), которые готовы на заказ произвести сборку кластера определённой вычислительной конфигурации, стоимость которого вполне доступна для промышленного сектора. Также существует возможность аренды необходимых вычислительных мощностей, в том числе с GPU, в облачных системах (например, Amazon).

Создание прикладной части последнего из перечисленных программных комплексов, реализующей математическое моделирование распространения сейсмических волн и их статистическое сопоставление с полевыми данными, было завершено несколько лет назад. Сегодня происходит расширение его функциональных возможностей. В то же время клиентское приложение в части визуализации, ввода и вывода данных, пред- и постобработки, анализа и интеграции данных пока находится в состоянии разработки.

Программное обеспечение для промышленного применения сводится не только к моделированию в составе прикладного программного комплекса, кроме собственно модельных расчётов, необходимо обеспечить эффективное выполнение следующих функций:

- управление данными большого объёма (сотни гигабайт и терабайт);
- постобработку модельных данных совместно с данными натуральных экспериментов;
- представление результатов пользователям в высококачественном графическом виде как с использованием 2D-, так и 3D-визуализации;
- коллективную работу с возможностью организации удалённого доступа.

В настоящий момент времени существуют программные технологии, позволяющие решить все эти задачи. Это технологии: организации распределённых вычислений MPI, OpenMP; высокопроизводительных вычислений на графических процессорах NVidia Cuda, OpenCL; реализации трёхмерной графики OpenGL и DirectX; распределённые файловые системы NFS, Samba и т. п. Однако названные технологии достаточно низкоуровневые, и при совместном использовании требуется решать вопросы их оптимальной интеграции при решении конкретных задач, отвлекаясь от создания прикладного кода для обеспечения производственного процесса. Важный момент также заключается в том, что используемые методики полноволновой сейсмологии сейчас находятся в стадии активного развития, следовательно, программное обеспечение должно быть легко модифицируемым под новую требуемую функциональность. Для этого должна быть предоставлена удобная и гибкая программная среда, которая позволяла бы производить изменения с минимальными временными и трудовыми затратами.

Таким образом, для создания полноценного промышленного программного комплекса необходим согласованный выбор программных и аппаратных технологий с минимальной сложностью в освоении и использовании, покрывающих все составные части прикладного программного обеспечения. Ниже дан обзор имеющихся программных и аппаратных технологий, которые могут использоваться для решения поставленной задачи — построения специализированного программного комплекса.

2. ИНТЕГРИРОВАННЫЕ ПРОГРАММНЫЕ КОМПЛЕКСЫ

В нефтегазовой отрасли, как и во многих других отраслях, существует большое количество программных продуктов, реализующих 2D- и 3D-визуализацию: Petrel, Landmark, GeoSoft Target, CoViz 4D. Ряд систем обеспечивает поддержку использования нескольких графических карт (Multi-GPU) — Paradigm VoxelGeo, TerraSpark-InsideEarth, f1A-GeoTeric. Некоторые из них, в частности, Petrel [Getting Started..., 2014; Ocean Wizards..., 2013], представляют собой расширяемые комплексные платформы: разработчик на их базе реализует только функциональную часть и связывает её с платформой согласованным образом. Однако эти решения ориентированы на архитектуру фон Неймана и не поддерживают архитектуру

туру распределённых систем. Кроме того, разрабатываемое программное обеспечение должно функционировать в рамках фирменных программных комплексов, стоимость которых крайне высока.

Существует программно-техническое решение компании NVidia — IndeX [NVIDIA IndeX..., 2014; Nienhaus et al., 2013]. Эта платформа позволяет осуществлять в реальном времени масштабируемую визуализацию объёмных и поверхностных данных за счёт распределенного рендеринга. Большое внимание уделяется интеграции с вычислительными кластерами, поэтому существует возможность объединения расчетных циклов с циклами визуализации. При визуализации не используются методы аппроксимации данных (Level of Detail), что гарантирует их достоверность и отсутствие артефактов, связанных с их искажениями. За управление распределением данных в системе, планированием и запуском вычислительных задач отвечает программная платформа Distributed Computing Environment (высокоуровневая альтернатива MPI) [DiCE TM 3.0..., 2009], которая позволяет разрабатывать приложения, масштабируемые на тысячи CPU-ядер и GPU-устройств. Разработку можно вести под различные операционные системы (Windows, Linux), но предлагается использовать только C++ API.

Программным решением, которое достаточно полно охватывает все аспекты разработки высокопроизводительного программного обеспечения для нефтегазовой отрасли, может рассматриваться HueSpace — программная платформа компании Hue AS (http://hue.no/sites/default/files/Hue_Brochure_US_web.pdf), [Walker, Buch, 2012]. Эта платформа позволяет оперировать большими объёмами сейсмических данных, производить их обработку и визуализацию. Поддерживается возможность организации вычислений на графических процессорах. Разработку можно вести под различные операционные системы (Windows, Linux) с использованием нескольких языков программирования (C++, .Net, Java). Существует пример использования системы в публичном облаке Amazon, что позволяет организовать ведение разработки программного обеспечения на арендуемых удалённых вычислительных мощностях.

Обе упомянутые платформы поддерживают расширение функциональности ядра, а также интеграцию с разрабатываемыми приложениями. Платформы считаются проприетарными, поэтому их описание в открытых источниках носит в основном информационный характер. В частности, несмотря на заявленное использование платформы HueSpace в программном продукте Roxar RMS, степень готовности последнего к массовому применению не ясна. Кроме того, платформы построены на базе программно-технических решений компании NVidia. Построенный нами комплекс базируется на графических ускорителях AMD — оптимальных сегодня по уровню аппаратных средств и программных технологий OpenCL.

Таким образом, разработку собственного программного продукта нельзя было осуществить на основе одной комплексной платформы — необходимы были выбор и интеграция программных решений, предназначенных для решения отдельных функциональных задач. Кроме того, был проведён анализ систем визуализации и управления хранением больших данных для выбора оптимального решения.

3. ВИЗУАЛИЗАЦИЯ МОДЕЛЬНЫХ ДАННЫХ

Наиболее подходящей для решения задачи визуализации модельных данных, получаемых на кластерном суперкомпьютере, представляется распределённая визуализация на тех же узлах и данных, которые использовались для моделирования. Задача распределённой обработки изображений в настоящий момент решается с помощью параллельного рендеринга графики, для чего используется большое количество программных продуктов. Традиционно рендеринг производился на стандартных центральных процессорах, но в настоящее время начинают активно использовать GPU-устройства, так как повышенная производительность особенно важна при применении технологий физически корректного рендеринга. Рассмотрим некоторые из таких технологий с возможностью распределённого функционирования.

NVIDIA Iray Server — новое программное решение от компании NVidia, которое должно обеспечить возможности распределённого рендеринга Iray на сетевых машинах. Этот программный продукт объединяет группу серверов в единую рендеринг-ферму и обеспечивает обработку изображений на всех процессорах и GPU-устройствах, доступных кластеру. Для кодирования видео используется H 264 кодек.

Equalizer — открытая платформа для создания и развёртывания многопоточных OpenGL-приложений. Программный пакет предоставляет C++ API для создания соответствующих приложений, которые могут далее исполняться в различном аппаратном окружении: от одиночных рабочих станций до крупных кластеров, в том числе с использованием возможностей Multi-GPU.

Пример другого подхода — программный пакет Chromium, который обеспечивает механизм встраивания параллельного рендеринга в существующие приложения. Встраивание обеспечивается за счёт осуществления перехвата OpenGL-вызовов, которые перенаправляются на кластер рендеринга. Несмотря на то, что подобный подход может быть менее эффективным [Stadt et al., 2003], на базе названного пакета создано несколько коммерческих пакетов и исследовательских кластеров рендеринга.

Распределённый рендеринг производится не только на локальных, аппаратно сильно связанных кластерах, но и в распределённых грид-системах, например, на основе платформы BOINC — на ней построены системы BURP и Electric Ship. Стоит отметить, что эти системы не претендуют на визуализацию в реальном времени.

Известны также работы по использованию парадигмы Map/Reduce экосистемы Hadoop для реализации различных алгоритмов распределённой визуализации данных [Howe et al., 2009; Stuart et al., 2010; Vo et al., 2011]. Это направление весьма перспективно, так как является альтернативным вариантом программной платформы визуализации, но в настоящий момент проводятся лишь научные исследования в этом направлении.

Среди существующих платформ нераспределённой визуализации можно выделить Open Inventor (http://www.openinventor.com/backoffice/wp-content/uploads/OpenInventor-OilGas-Mining-Geosciences_FEI-VSG_BR.pdf; http://www.openinventor.com/backoffice/wp-content/uploads/OpenInventor-RemoteViz_FEI-VSG_DS.pdf) от компании FEI VSG — инструмент для создания программных продуктов со сложной 2D- и 3D-графикой. Платформа содержит широкий набор средств и методов визуализации и анализа данных для нефтегазовой промышленности и активно использует возможности технологии GPU для высокопроизводительной визуализации, а также имеет возможности работы с большими объё-

ёмами данных. Поддерживаются облачные, настольные и мобильные варианты. Разработка может вестись на языках программирования C++, C# (.NET) или Java. Данный инструмент предоставляет большие возможности для разработки графического программного обеспечения, но в нем отсутствуют средства для организации распределённых вычислений. Известно применение этой программной платформы в разработке программных решений компании Schlumberger [Yrke, Neck, 2015].

Компания INT создала платформу визуализации и разработки программного обеспечения INTViewer (<http://www.int.com/images/int/intviewer/doc/INTViewer5.1.pdf>) для анализа сейсмических данных и контроля их качества. В основе системы лежит платформа разработки Java Swing приложений — Netbeans Rich Client Platform. Разработка может вестись на языках программирования Java и Python, возможна интеграция с Matlab. Обеспечена поддержка различных операционных систем (Windows, Mac, Linux, UNIX). Предоставляются инструменты для разработки веб-приложений. Прodeкларирована возможность работы с данными любого размера, однако техническая реализация не приводится, также не ясна возможность использования GPU и распределённой обработки данных.

Известны также системы, ориентированные на визуализацию в предметно-ограниченной области. Особенно востребована высокопроизводительная визуализация научных данных. Здесь технические сложности связаны с большим объёмом данных (терабайты), которые требуется графически отобразить пользователю. Известные программные решения в основном представляют собой полноценные программные продукты с инструментами визуализации и анализа данных и опциональной возможностью расширения функциональности за счёт системы расширений (плагинов). Одна из наиболее известных систем такого рода — программный продукт Paraview. Он поддерживает клиент-серверную архитектуру для организации удалённой визуализации массивов данных и использует метод изменения уровня детализации для поддержки визуализации больших объёмов данных в интерактивном режиме. На его основе в университете Иллинойса разрабатывается программное решение RVA (Reservoir Visualization and Analysis) [RVA..., 2015] для 3D-визуализации данных в области моделирования нефтяных залежей и анализа геоданных. Отметим наличие дополнительных компонент ParaviewWeb, которые позволяют отображать данные с использованием технологий HTML 5.0 и WebGL в браузере.

Среди других аналогичных программных пакетов, предназначенных для параллельного рендеринга и визуализации больших данных, выделим программные пакеты Vaa3D, VisIt и Fiji. Исследователи реализуют на их основе программные решения для отображения больших данных в различных областях науки, например, BigDataVisualizer [Pietzsch et al., 2015] (плагин к Fiji) — инструмент для просмотра терабайтных последовательностей 3D-изображений, полученных с помощью световой микроскопии; Vaa3D-TeraFly [Bria et al., 2015] (плагин к Vaa3D) — для просмотра и анализа больших данных биологических исследований [Hassan et al., 2012].

Активно развивается такое направление визуализации, как «виртуальная реальность». Примером служит одно из интереснейших направлений в моделировании виртуальных миров — системы CAVE (Cave Automatic Virtual Environment). Среди программно-аппаратных комплексов, наиболее приближенных к рассматриваемой нами тематике, отметим решение компании TechViz [<http://www.techviz.net/techviz-xl>], которое предоставляет «драйвер» (TechViz XL) и программную

библиотеку (TVZLib API), позволяющие прозрачно транслировать 3D OpenGL-приложения в «виртуальное» окружение. Преимуществом этого решения является декларируемая возможность прозрачного использования (интеграции) с большим списком существующего программного обеспечения как из области 3D-визуализации, так и из нефтегазовой области. Также можно отметить решение MiddleVR от компании VisBox — в нём реализована интеграция с программным пакетом Unity создания 3D-приложений и игр (в том числе, в веб-сайтах).

Естественным представлением сеточных моделей является воксельная модель графических данных [Kaufman et al., 1993]. Эти модели хорошо подходят для моделирования непрерывных сред и полей значений, что характерно, например, для сейсмических и ряда медицинских данных. Тем не менее, воксельная графика, наиболее естественно подходящая для представления данных численного моделирования, находится лишь на ранней стадии развития. Известны отдельные эксперименты и закрытые решения, в том числе в области геологии и медицины, что показывает практическую применимость этой модели для представления сложных пространственных данных, однако отчуждаемых решений, которые могли бы послужить базой для прикладной системы, пока нет.

В медицине, как и в геологии, существенное значение для анализа и интерпретации данных играет опыт исследователя, и требуется качественная и детальная визуализация объекта наблюдения. Созданное программное обеспечение использует как базовые библиотеки визуализации, например, VTK, так и более специализированные инструменты — библиотеки MeVisLab, MIPAV и другие [Андреева, Маркина, 2013; Caban et al., 2007]. Специализированные библиотеки (ITK) применяются для обработки и предварительной подготовки изображений. Воксельная графика достаточно давно используется в компьютерных играх, но, поскольку этот альтернативный метод визуализации изначально более ресурсоёмкий, чем использование стандартных полигональных моделей, в других областях его распространение было ограничено. Однако в настоящий момент с появлением новых высокопроизводительных средств, например, GPU [Panteleev, 2014], воксельная графика начинает активно применяться на практике, например, в графической платформе компании ОТОУ и ряде проприетарных программных решений для анализа геологических и геофизических данных, в частности, в системе VoxelGeo компании Paradigm. Рендер Octane Render Cloud от компании ОТОУ интересен тем, что у него существует облачная бета-версия программного продукта. Возможно функционирование как в публичном облаке Amazon, так и на собственном GPU-кластере компании. Для трансляции видео используется проприетарный ORBX видекодек.

Практически все перечисленные выше системы рассчитаны на визуализацию традиционных графических данных, представленных полигональной моделью. Результаты же численного моделирования имеют принципиально другую природу — это статические или динамические значения физических величин в узлах сетки, которые необходимо перевести в графические величины (оптическая плотность, цвет) и визуализировать. Для использования полигональной визуализации требовались динамическое преобразование значений параметров моделирования в узлах сетки в набор полигонов и выделение под них отдельного подмножества узлов под рендер-ферму. Такое решение неоптимально по затратам на преобразование данных.

4. УДАЛЁННЫЙ ДОСТУП

Вследствие специфики деятельности нефтегазовых компаний, в частности, её распределённого характера, предлагаемые программные комплексы для обработки, анализа и визуализации производственных данных должны иметь возможность функционирования в удалённом режиме. Кроме того, процедура анализа информации и принятия решений характеризуется коллективной работой специалистов, а это также накладывает требования на организацию удалённого доступа.

Классическим вариантом создания удалённого сервера доступа к данным можно считать виртуализацию рабочих столов с помощью специализированного программного обеспечения. Это направление развивается давно, хорошо известно и поддерживается ведущими мировыми вендорами (VMWare, Citrix, Microsoft). Однако долгое время такой подход не давал возможности работы с «тяжелой» 3D-графикой и был предназначен для работы с офисными программами. Ситуация изменилась после появления аппаратных средств, обеспечивающих ускорение обработки графики в виртуализированной среде. Существуют различные варианты данного механизма: GPU pass-through (один GPU на одну виртуальную машину), Shared GPU (программная виртуализация GPU), Virtual GPU (аппаратная виртуализация GPU). Наиболее эффективным и быстродействующим считается механизм vGPU. Эти возможности предоставляются видеокартами NVidia GRID K1 и K2 [Currid, 2014], а также поддерживающими технологию AMD Multiuser GPU серверные карты FirePro S-series. Данные устройства обладают возможностью настройки различных профилей виртуальных GPU в зависимости от требований используемого программного обеспечения и необходимого количества пользователей (до 16). Современные сервера способны организовать высокую плотность компоновки подобных карт (несколько видеокарт в 1U) и, значит, высокую аппаратную мощность.

Техническое решение для предоставления удалённого доступа к серверам, в том числе к системам виртуализации, предлагает компания Teradici. Разработанная ею технология PCoIP позволяет осуществлять удалённый доступ к рабочим столам. Протокол PCoIP предусматривает сжатие, шифрование и кодирование информации об экранном буфере, обеспечивает передачу PCoIP-устройствам только данных об изменившихся пикселях. Поддерживаются передача изображения высокой чёткости и более высокого разрешения, а также частота кадров, пригодная для работы с трёхмерной графикой. Поддержка данного решения производится в настоящий момент и вендорами — производителями видеокарт, например, такая возможность реализуется с помощью линейки продукты AMD FirePro R5000 Remote Graphics. Известно, что на производительность системы при визуализации данных оказывает влияние центральный процессор, который занимается кодированием видеопотока для отправки его клиенту. Это добавляет существенную нагрузку на систему. В таком случае дополнительное использование устройств Teradici позволяет значительно увеличить производительность.

Другим вариантом организации удалённого доступа можно считать использование технического подхода, применяемого в интернет-сервисах предоставления и трансляции потокового видео и аудио (YouTube, Netflix и т. д.) на устройства конечного пользователя. Такой подход стал также использоваться в сервисах цифровой дистрибуции компьютерных игр и известен как Gaming as a Service или Games on Demand [Kushner, 2009; Ojala, Tyrväinen, 2011; Ross, 2009]. Технически игровой пакет размещается и исполняется (обрабатывается)

на стороне удалённого сервера, а пользователям поставляются лишь результирующие данные (видео и звук) через интернет-соединение. Это позволяет производителям игр практически не оглядываться на аппаратные возможности пользовательских машин. С другой стороны, проявляется большая зависимость от качества и скорости соединения с игровыми серверами. Описанное направление коррелирует с концепцией создания и реализации интерактивной трёхмерной компьютерной графики в реальном времени, которую развивает, например, компания AMD в рамках проекта Cinema 2.0. Её суть состоит в создании полностью интерактивных компьютерных игр и других приложений с кинореалистичной графикой. Визуализация данных в облаках, облачный рендеринг и облачные игры в настоящий момент времени активно развиваются, и известно достаточно большое количество программно-технических решений и соответствующих вендоров (OTOY, G-cluster, PlayStation Now). Игровые сервисы могут предоставлять как персональные компьютеры, так и мобильные устройства, а также игровые консоли. Контент может распространяться как через общедоступные интернет-каналы, так и через телевизионные кабельные и IPTV-сети.

Для организации технической возможности реализации удалённой обработки и трансляции качественной видеографики требуются соответствующие аппаратные решения. В настоящий момент необходимое оборудование предоставляют компании NVidia и AMD: у NVidia это линейка графических карт GRID K340 и GRID K520, построенных на архитектуре Kepler; а у AMD — серия графических карт Radeon Sky на архитектуре Graphics Core Next (GCN). С использованием соответствующих технологий NVidia GRID и AMD RapidFire реализуется возможность визуализации и потоковой передачи игр на различные вычислительные устройства. В основном названные технологии не направлены на распределённый рендеринг, то есть на обработку контента несколькими устройствами, наоборот, графические ускорители осуществляют параллельную обработку нескольких игровых потоков (до 6 в HD(720p) качестве для Radeon Sky).

К дополнительному преимуществу описанных технологий можно отнести повышенную безопасность данных, что особенно важно в коммерческих программных решениях для нефтегазового сектора: пользователю передаётся только видеоинформация, утечка исходных данных минимизирована.

Итак, технологии организации удалённого доступа давно используются в промышленности, что позволяет говорить о высоком уровне их зрелости.

5. УПРАВЛЕНИЕ БОЛЬШИМИ ДАННЫМИ

Важным вопросом считается управление большими данными, возникающими как в процессе моделирования, так и в натурных экспериментах. Полевые данные с датчиков при длительном мониторинге могут достигать от нескольких до сотен терабайт. Передача таких данных внутри системы занимает несколько минут даже по сетям Infiniband, внешняя архивация практически невозможна. «Узким местом» становится интерфейс к дисковой подсистеме, со скоростью поточного считывания даже в RAID-системах — не более 1–2 Гб/с. Очевидный механизм разрешения проблемы — распределённое хранение данных, при котором скорость доступа ко всему файлу, например, при копировании, увеличивается пропорционально количеству устройств хранения. Другой аспект доступа — необходимость индексного доступа к части файла для визуализации или анализа, чтобы избежать последовательного поиска в файле. Учитывая сложную структуру сеточных моделей, можно утверждать, что это — нетривиальная задача.

Отдельная проблема — обеспечение надёжности хранения. Данные моделирования требуется резервировать внутри системы, однако степень избыточности данных необходимо регулировать. Тотальное дублирование данных представляется избыточной мерой для легко воспроизводимых данных собственно моделирования, тогда как полевые данные, потеря которых невосполнима, требуется резервировать многократно. Таким образом, данные внутри системы необходимо подвергать управляемому резервированию.

В настоящий момент времени работа с большими массивами данных происходит в рамках развиваемой парадигмы BigData. Наиболее известная и распространённая программная технология в этой области — Hadoop [Kerzner, Maniyam, 2013]. Эта платформа обладает широко развитой экосистемой, активно идёт процесс коммерциализации технологии — ряд компаний строит свой бизнес целиком на создании коммерческих дистрибутивов Hadoop и услуг по технической поддержке. Изначально технология Hadoop предназначалась в основном для обеспечения высоконагруженных процессов, связанных с функционированием веб-платформ, но происходит внедрение в другие отрасли, в том числе, в нефтегазовую. Последняя очень серьёзно относится к перспективам использования названной технологии [Febowitz, 2012; Holdaway, 2014; Priyadarshy, 2015]. Крупные программные вендоры предлагают свои решения для хранения, интеграции и анализа данных (IBM BigData, SAP BigData и т. д.). Но эти программные продукты в основном нацелены на анализ данных. Известны и гибридные системы [Mitchell et al., 2011], в которых Hadoop используется для хранения данных, а программный пакет Paraview осуществляет их визуализацию.

В области обработки данных сейсмических исследований также ведутся работы по изучению эффективности использования Hadoop. Например, Seismic Hadoop позволяет производить обработку сейсмических данных с помощью пакета Seismic Unix в Hadoop-окружении от компании Cloudera. Также рассматривается [Li et al., 2015] возможность хранения сейсмических данных (модифицированный SEG-Y формат) как в стандартных распределённых хранилищах (Fast DFS), так и в хранилищах собственной разработки, также проводятся сравнения с HDFS. Но пока проанализирован доступ к файлам небольших размеров. Однако необходимость и эффективность использования Hadoop-платформы в качестве управления данными для названной системы не очевидна. Во многом это связано с отличием в парадигме обработки и хранения данных, для работы с которыми система предназначалась изначально, от парадигм, используемых в научных и промышленных вычислениях. Например, модельные вычисления в основном характеризуются интенсивными локальными расчётами, генерирующими большие объёмы данных текущего состояния на каждой итерации. Сохранение состояния между шагами Map/Reduce требует записи в файловую систему, что является очень времязатратной операцией. Также используемые алгоритмы требуют прямого взаимодействия между узлами, что не поддерживается инфраструктурой Map/Reduce. В итоге некоторые исследователи констатируют, что использование платформы Hadoop для научных вычислений в 95% случаев может привести к замедлению производительности на 1–2 порядка [Stonebraker, Kerper, 2012].

Используемая в файловой системе Hadoop HDFS байтовая потоковая модель данных эффективно функционирует на массиве неструктурированной информации текстового формата, но не оптимальна при работе с высокоструктурированными бинарными форматами файлов, в которых обычно представляются данные научных исследований и производственных операций. Эффективная работа HDFS возможна, когда реализуется быстрое поблочное чтение файлов, с возможностью

осуществления параллельных операций над ними. Если же одновременно поблочное чтение и обработка несовместимы, что характерно для больших и сильно структурированных файлов, то будет происходить резкое падение производительности. Особенностью массива модельных данных как раз является то, что здесь имеют дело не только с большой совокупностью небольших файлов, но и с большой совокупностью файлов гигантских размеров. Системы, по умолчанию рассчитанные на хранение файлов средних (Hadoop) или небольших размеров, могут организовать хранение и больших файлов, но не оптимальным способом (происходит разбишка на блоки без учёта внутренней структуры), что в общем случае снизит скорость их обработки. Для Hadoop декларирована разработка возможности использования технологии Авто-контейнеров для обеспечения оптимизированного хранения HDF5-файлов (один из распространённых стандартов хранения данных в научных исследованиях) небольших и средних размеров [Heber et al., 2014]. Большие файлы (больше размера блока) могут быть разбиты на временные серии для различных локаций и храниться в таком виде, что обеспечит быстрое представление информации потребителю. В случае геологических данных или данных моделирования, которые представляют собой временные серии, это выглядит хорошим вариантом решения проблемы, хотя и вносит дополнительное преобразование данных, однако описанная техническая возможность ещё не реализована. Другое возможное решение видится в создании Hadoop-плагинов — SciHadoop [Buck et al., 2011], который позволяет организовывать логические запросы к моделям, данные которых организованы в виде массивов. Для хранения данных используется формат NetCDF-3, основой решения является функция преобразования логических координат элементов данных в смещение в физическом файле. Существует и другая возможность, например, использование HDF5 Virtual Object Layer. Таким образом, поиск технического решения ведётся, но пока варианта для промышленного использования не предложено.

В целом необходимо отметить, что вопрос оптимального взаимодействия с данными в HDF5-формате возникает и при использовании других распределённых систем хранения, например, известны экспериментальные работы по увеличению в десятки раз скорости работы с такими файлами в Lustre File Systems [Howison, 2010].

Ещё одна проблема — избыточность данных. До Hadoop ещё не были распространены и широко доступны быстрые сети на основе 10G Ethernet и Infiniband, это обусловило ряд архитектурных особенностей системы. В частности, данные хранятся в основном там, где генерируются (2 экземпляра копии), и лишь одна копия хранится на удалённых серверах. Сейчас же скорость современных сетей превышает скорость записи на группу дисков, значит, можно преимущественно реализовывать удалённое хранение и хранить данные там, где они должны быть использованы. Это позволяет также использовать более сложные схемы хранения, но ускорять при этом запись и повышать надёжность. Примером таких систем нового поколения служит Quantcast File System, в которой используется кодирование Рида–Соломона для хранения данных одновременно в 9 местах и восстановления по любым 6 блокам. Это повышает надёжность данных, а избыточность хранения составляет только 50 %. Однако в этом случае разбишка на блоки производится без учёта логической структуры файла, что может привести к осязаемому снижению производительности.

Альтернативой Hadoop могут стать программные решения, основанные на объектно-ориентированных распределённых файловых системах и позволяющие строить сравнительно недорогие, масштабируемые и самовосстанавливаемые хранилища данных. В настоящий момент они используются для хранения данных

такими компаниями, как Facebook, Spotify, Dropbox. Примерами таких систем служат Swift, Ceph Storage System.

Первоначально инструменты Hadoop базировались на файловой системе HDFS, однако работа с данными на уровне файловой системы не всегда удобна. Для выполнения многих высокоуровневых функций анализа и обработки данных предпочтительным является использование более сложных программных комплексов и баз данных. Отметим, что в данном направлении происходит развитие экосистем Hadoop и Big Data, т. е. от пользователей скрываются сложности функционирования Map/Reduce и HDFS за счёт предоставления более высокоуровневых инструментов, таких как HBase, Hive, Pig и Impala.

В настоящее время существующие распространённые базы данных не очень хорошо подходят для модельных данных, так как последние в основном представляют собой многомерные массивы огромных размеров. Существующие технические возможности традиционных баз данных обеспечивают только хранение (например, с помощью BLOB-объектов), но не гарантируют оптимального доступа к данным для анализа и не предоставляют средства для гибкого оперирования данными. Такие возможности могут быть предоставлены системами управления базами данных на основе массивов (array-based DBMS), которые оптимизируют хранение многомерных данных и обеспечивают наличие стандартного языка (QL) работы с ними. К таким системам относятся SciDB [Cudrè-Mauroux et al., 2009] и Rasdaman [Baumann et al., 1999]. Отметим, что SciDB изначально разрабатывалась как система для хранения и обработки данных научных исследований и основана на парадигме хранения данных в виде многомерных вложенных массивов. Однако пока такие системы управления данными не являются распространёнными.

Как мы отмечали, решение задачи управления большими данными результатов моделирования требует одновременно индексного доступа, распределённого хранения и оптимального резервирования. Комплексная система управления, удовлетворяющая одновременно всем трём требованиям, отсутствует. Имеющиеся в настоящее время доступные системы управления данными ориентированы на выполнение какой-то одной из этих функций, и их интеграция требует неприемлемых накладных расходов.

ЗАКЛЮЧЕНИЕ

В связи с отсутствием на рынке необходимой комплексной программной платформы для прикладных систем на базе численного моделирования, нами построен программно-аппаратный комплекс, который во многом является компромиссным решением и внутренним программно-технологическим полигоном. С нашей точки зрения, в этом комплексе достаточно эффективно решены задачи построения аппаратного слоя, слоя организации высокопроизводительных вычислений, а также обеспечена возможность удалённого доступа пользователей и трансляции видеоданных с помощью системы виртуализации. Подсистема хранения данных и распределённый рендеринг интерактивных 3D-изображений — части комплекса, которые требующие дальнейшей разработки.

В настоящий момент эффективность подсистемы хранения обеспечивается с помощью аппаратной платформы. Для этой цели используется RAID-массив, состоящий из высокоскоростных дисков, а также организован SSD-кеш. В качестве файловой подсистемы выбрана одна из самых простых, но эффективных, — NFS, что с учётом централизованного дискового массива и организованной

Infiniband-сети позволяет осуществлять достаточно быстрое сохранение данных вычислительными узлами. Это также обеспечивается сравнительно небольшим количеством вычислительных процессов (не более 24), которые осуществляют запись информации. Однако при дальнейшей модернизации комплекса и росте количества процессов эффективность ввода/вывода начнёт снижаться, и станет необходимо рассмотреть возможность использования распределённой системы хранения (файловой или СУБД).

Ближайшей целью совершенствования разработанного комплекса будет развитие слоя распределённой визуализации данных. В настоящий момент визуализация данных происходит с использованием возможностей карт NVidia K2 в виртуальной среде Citrix Desktop. Однако по имеющимся оценкам скорость визуализации будет недостаточна уже в ближайшем будущем при отображении результатов геофизических исследований и численного моделирования больших месторождений.

Таким образом, создание специализированной комплексной высокопроизводительной программной платформы является актуальной задачей. Её решение позволило бы уменьшить сложность системного и прикладного программного обеспечения, обеспечить масштабируемость в зависимости от предъявляемых требований и адаптацию под будущие программно-технические решения и в итоге снизить стоимость внедрения и эксплуатации суперкомпьютерных комплексов в промышленности.

С учётом больших объёмов используемых данных оптимальным представляется построение платформы на основе единой модели данных для визуализации, расчётов и хранения для сокращения затрат вычислительных ресурсов на преобразования данных в процессе работы и минимизации занимаемой памяти.

Доступная программная платформа для построения прикладных систем с использованием суперкомпьютерного численного моделирования позволила бы увеличить количество научно-технологических стартапов, реализующих новые технические подходы в промышленности, причём создаваемые ими решения были бы совместимы на программном уровне и могли бы быть интегрированы в едином бизнес-процессе предприятий.

ЛИТЕРАТУРА

- [Андреева, Маркина, 2013] *Андреева А. Д., Маркина С. Э.* Обзор программ для визуализации медицинских данных // Молодой учёный, 2013. № 3. С. 512–516.
- [Бетелин и др., 2007] *Бетелин В. Б., Велихов Е. П., Кушниренко А. Г.* Массовые суперкомпьютерные технологии — основа конкурентоспособности национальной экономики в XXI веке // Информационные технологии и вычислительные системы, 2007. № 2. С. 3–10. URL: http://www.jitcs.ru/index.php?option=com_content&view=article&id=17.
- [Шабалин и др., 2013] *Шабалин Н. Я., Рыжов В. А., Биряльцев Е. В.* Пассивная низкочастотная сейсморазведка — мифы и реальность // Приборы и системы разведочной геофизики, 2013. № 2. С. 46–53.
- [Baumann et al., 1999] *Baumann P. et al.* Spatio-temporal retrieval with RasDaMan // VLDB, 1999. P. 746–749.
- [Bria et al., 2015] *Bria A., Iannello G., Peng H.* An open-source VAA3D plugin for real-time 3d visualization of terabyte-sized volumetric images // Biomedical Imaging (ISBI), 2015 IEEE 12th Intern. Symp. 16–19 April 2015. P. 520–523. URL: http://home.penglab.com/papersall/docpdf/2015_ISBI_Terafly.pdf.

- [Buck et al., 2011] *Buck J. B.* et al. Scihadoop: array-based query processing in Hadoop // Proc. 2011 Intern. Conf. High Performance Computing, Networking, Storage and Analysis. ACM, 2011. P. 66.
- [Caban et al., 2007] *Caban J. J., Joshi A., Nagy P.* Rapid development of medical imaging tools with open-source libraries // J. Digital Imaging. 2007. V. 20. No. 1. P. 83–93.
- [Cudré-Mauroux et al., 2009] *Cudré-Mauroux P.* et al. A demonstration of SciDB: a science-oriented DBMS // Proc. of the VLDB Endowment. 2009. V. 2. No. 2. P. 1534–1537.
- [Currid, 2014] *Currid A.* Delivering high-performance remote graphics with NVIDIA GRID virtual GPU // SIGGRAPH. 2014. URL: <http://on-demand.gputechconf.com/siggraph/2014/presentation/SG4119-Delivering-High-Performance-Remote-Graphics-NVIDIA-GRID-Virtual-GPU.pdf>.
- [Da Costa et al., 2015] *Da Costa G.* et al. Exascale machines require new programming paradigms and runtimes // Supercomputing Frontiers and Innovations. 2015. V. 2. P. 6–27.
- [DiCE TM 3.0..., 2009] DiCE TM 3.0 Functional Overview. White Paper, mental images GmbH. 2009. URL: http://www.nvidia-arc.com/fileadmin/user_upload/PDF/DiCE_White_Paper.pdf.
- [Febowitz, 2012] *Febowitz J.* The big deal about big data in upstream oil and gas // Paper and Presentation, IDC Energy Insights. 2012. URL: <https://www.hds.com/assets/pdf/the-big-deal-about-big-data-in-upstream-oil-and-gas.pdf>.
- [Getting Started..., 2014] Getting Started with Ocean for Geoscientists and Software Developers. Published by Schlumberger Information Solutions, San Felipe, Houston Texas. 2014. URL: <https://www.ocean.slb.com/~media/files/developer/getting%20started/gettingstartedwithocean-2014-1.pdf>.
- [Haffinger, 2013] *Haffinger P. R.* Seismic broadband full waveform inversion by shot/receiver refocusing. TU Delft, Delft University of Technology, 2013. URL: <http://repository.tudelft.nl/view/ir/uuid%3Ad2d8d264-5037-4573-8418-a079afa8d1e7>.
- [Hassan et al., 2012] *Hassan A. H., Fluke C. J., Barnes D. G.* A distributed GPU-based framework for real-time 3d volume rendering of large astronomical data cubes // Publications of the Astronomical Society of Australia. 2012. V. 29. No. 03. P. 340–351.
- [Heber et al., 2014] *Heber G., Folk M., Koziol T.* Big HDF FAQs. The HDF Group. 2014. http://www.hdfgroup.org/pubs/papers/Big_HDF_FAQs.pdf.
- [Holdaway, 2014] *Holdaway K.* Harness Oil and gas big data with analytics: optimize exploration and production with data driven models. John Wiley and Sons, 2014. URL: http://www.sas.com/storefront/aux/en/spbigdataoilgas/66509_excerpt.pdf.
- [Howe et al., 2009] *Howe B.* et al. Query-driven visualization in the cloud with MapReduce // Proc. of the Fourth Annual Workshop on Ultrascale Visualization. 2009. URL: <http://citeserx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.9186&rep=rep1&type=pdf>.
- [Howison, 2010] *Howison M.* Tuning HDF5 for Lustre file systems // Workshop on Interfaces and Abstractions for Scientific Data Storage (IASDS10). Heraklion, Crete, Greece, September 24, 2010. URL: https://www.hdfgroup.org/pubs/papers/howison_hdf5_lustre_iasds2010.pdf.
- [Kaufman et al., 1993] *Kaufman A., Cohen D., Yagel R.* Volume graphics // Computer. 1993. V. 26. No. 7. P. 51–64.
- [Kerzner, Maniyam, 2013] *Kerzner S. M. M., Maniyam S.* Hadoop Illuminated // Hadoop Illuminated LLC. 2013. URL: http://hadoopilluminated.com/hadoop_illuminated/hadoop-illuminated.pdf.
- [Kushner, 2009] *Kushner D.* Game developers see promise in cloud computing, but some are skeptical // The IEEE Spectrum March. 2009. URL: <http://spectrum.ieee.org/computing/networks/game-developers-see-promise-in-cloud-computing-but-some-are-skeptical/1>.
- [Li et al., 2015] *Li J., Pan C., Lu M.* A seismic data processing system based on fast distributed file system // Intern. J. Computers and Technology. 2015. V. 14. No. 5. URL: http://cir-world.com/journals/index.php/ijct/article/view/4045/pdf_586.

- [Mitchell et al., 2011] *Mitchell C., Ahrens J., Wang J.* Visio: enabling interactive visualization of ultra-scale, time series data via high-bandwidth distributed i/o systems // Parallel and Distributed Processing Symp. (IPDPS). 2011 IEEE Intern. IEEE, 2011. P. 68–79.
- [Nienhaus et al., 2013] *Nienhaus M., Mensmann J., Radig S., Thamm T.* NVIDIA indeX — leveraging GPU compute for interactive large-scale data visualization // GPU Technology Conference. 2013. URL: <http://on-demand.gputechconf.com/gtc/2013/presentations/S3415-indeX-Interactive-Data-Visualization.pdf>.
- [NVIDIA IndeX..., 2014] NVIDIA IndeX, version 1.2. Whitepaper, NVIDIA Advanced Rendering Center. 2014. URL: http://www.nvidia-arc.com/fileadmin/user_upload/index_2014/documents/nvidia_index_whitepaper.140505.A4.pdf.
- [Ocean Wizards..., 2013] Ocean Wizards and Developer's Tools in Visual Studio for Geoscientists and Software Developers. Published by Schlumberger Information Solutions, San Felipe, Houston Texas. 2013. URL: <https://www.ocean.slb.com/~media/files/developer/getting%20started/oceanwizardsanddeveloperstools2013.pdf>.
- [Ojala, Tyrväinen, 2011] *Ojala A., Tyrväinen P.* Developing cloud business models: A case study on cloud gaming // IEEE Software. 2011. V. 28. No. 4. P. 42–47. URL: http://www.computer.org/cms/Computer.org/ComputingNow/homepage/2011/0711/T_SW_DevelopingCloudBusinessModels.pdf.
- [Panteleev, 2014] *Panteleev A.* Practical real-time voxel-based global illumination for current GPUs // SIGGRAPH. 2014. URL: <http://on-demand.gputechconf.com/siggraph/2014/presentation/SG4114-Practical-Real-Time-Voxel-Based-Global-Illumination-Current-GPUs.pdf>.
- [Pietzsch et al., 2015] *Pietzsch T.* et al. BigDataViewer: visualization and processing for large image data sets // Nature Methods. 2015. V. 12. No. 6. P. 481–483.
- [Priyadarshy, 2015] *Priyadarshy S.* (2015) The 7 pillars of Big Data // Petroleum Review. 2015. URL: https://www.landmark.solutions/Portals/0/LMSDocs/Whitepapers/The_7_pillars_of_Big_Data_Whitepaper.pdf.
- [Ross, 2009] *Ross P. E.* Cloud computing's killer app: gaming // Spectrum, IEEE. 2009. V. 46. No. 3. P. 14–14.
- [RVA..., 2015] RVA: A Plugin to ParaView 3.14 for Improved Reservoir Visualization and Analysis (2015): User manual version 2.5.2, September. 2015. URL: http://rva.cs.illinois.edu/downloads-v1/RVAmmanual_v252.pdf.
- [Stadt et al., 2003] *Stadt O. G.* et al. A survey and performance analysis of software platforms for interactive cluster-based multi-screen rendering // Proc. Workshop on Virtual environments 2003. ACM. 2003. P. 261–270.
- [Stonebraker, Kepner, 2012] *Stonebraker M., Kepner J.* Possible Hadoop trajectories. Communications of the ACM, Blog. 2012. URL: <http://cacm.acm.org/blogs/blog-cacm/149074-possible-hadoop-trajectories/fulltext>.
- [Stuart et al., 2010] *Stuart J. A.* et al. Multi-GPU volume rendering using MapReduce // Proc. of the 19th ACM Int. Symposium on High Performance Distributed Computing. ACM. 2010. P. 841–848.
- [Vo et al., 2011] *Vo H. T.* et al. Parallel visualization on large clusters using MapReduce // Large Data Analysis and Visualization (LDAV). 2011 IEEE Symp. 2011. P. 81–88.
- [Walker, Buch, 2012] *Walker K., Buch D.* Integrated GPU acceleration with real time visualization of terabyte data. NVidia GTC, Hue AS/Hue Technology North America. 2012. URL: http://hue.no/sites/default/files/S0436-Integrated-GPU-Acceleration-RT-Visualization-of-Terabyte-Data_0.pdf.
- [Yrke, Heck, 2015] *Yrke O., Heck M.* How Schlumberger leverages NVIDIA GPUs using the open inventor toolkit // GTC. 2015. URL: <http://on-demand.gputechconf.com/gtc/2015/presentation/S5329-Michael-Heck-Oyvind-Yrke.pdf>.

HPC CLUSTER DEVELOPMENT AND OPERATION EXPERIENCE FOR SOLVING THE INVERSE PROBLEMS OF SEISMIC PROSPECTING

*E. V. Biryaltsev*¹, *M. R. Galimov*¹, *D. E. Demidov*², *A. M. Elizarov*^{3,2}

¹ "Gradient Technology", Kazan

² Kazan Branch of Joint Supercomputer Center of the Russian Academy of Sciences

³ Lobachevskii Institute of Mathematics and Mechanics of Kazan Federal University (IMM KFU)

The experience of creating an integrated hardware and software solutions for high-performance computing in solving the problems of modern geophysical techniques based on full-wave inversion is described. Problems creating a mass of high-performance software systems for a wide use in industry are discussed. Review of existing software and hardware solutions in this area is given.

Keywords: graphics general-purpose computing, numerical simulation, distributed computing, full-wave inversion, GPU cluster, geophysical surveys, high-performance computing platform.

Biryaltsev Evgenij Vasil'evich — general director, PhD, igenbir@ya.ru

Galimov Marat Rasifovich — deputy general director, PhD, glmvmrt@gmail.com

Demidov Denis Evgen'evich — senior researcher, PhD, denis.demidov@gmail.com

Elizarov Alexander Mikhajlovich — doctor of physical and mathematical sciences, professor IMM KFU, head of the Kazan Branch of Joint Supercomputer Center of RAS, amelizarov@gmail.com

ТРЕХМЕРНАЯ ТЕНЗОРНАЯ МАТЕМАТИКА ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ В ГИДРОМЕХАНИКЕ

А. В. Богданов, А. Б. Дегтярев, В. Н. Храмушин

Санкт-Петербургский государственный университет (СПбГУ)

Архитектура цифрового вычислительного комплекса определяет техническую основу унифицированного математического языка для строгого арифметико-логического описания явлений и законов механики сплошных сред для прикладных задач гидромеханики и физической теории поля. Совершенствование современных компьютеров стоит на пути вовлечения вычислительных операций в весьма вольные формализации множества локально-связных цифровых геометрических объектов, независимо существующих и трансформирующихся в окружении аналогичных виртуальных сущностей, и по совокупности моделирующих реальные взаимодействия смежных частиц сплошной среды. Глубокое распараллеливание вычислительных процессов служит возрождению на новом техническом уровне прикладного функционального программирования, обеспечивающего сквозной контроль вычислительных процессов с автоматическим применением разномасштабных гибридных схем и адаптивных математических моделей для достоверного воспроизведения фундаментальных законов физики и механики сплошных сред. Тензорная формализация числовых объектов и вычислительных операций служит пространственной интерполяции параметров реологического состояния и законов гидромеханики в качестве математических моделей в локальных координатах элементарных числовых ячеек — крупных частиц жидкости.

Работа частично подддержана Российским фондом фундаментальных исследований (проект № 13-07-00747-а) и Санкт-Петербургским государственным университетом (проект № 9.38.674.2013). Исследования проведены с использованием оборудования ресурсного центра «Вычислительный центр СПбГУ».

Ключевые слова: вычислительный эксперимент, тензорная математика, гидромеханика.

Основные обозначения:

$\hat{r} = r_{ij}$ [м ³]	тензор формы крупной частицы жидкости;
$\overset{<}{v} = \hat{v} \cdot \overset{\vee}{r}$ [с ⁻¹]	тензор конвективных скоростей;
$\overset{\vee}{\rho} = \rho^{ik}$ [кг/м ³]	анизотропная динамическая плотность жидкости;
$\overset{>}{m} = \overset{\vee}{\rho} \cdot \hat{r}$ [кг]	объёмная (инерционная) масса виртуальной ячейки (частицы);
\vec{f} [Н]	массовые силы, распределённые по объёму частицы жидкости;
\hat{f} [Н·м ²]	тензор поверхностных напряжений на гранях ячейки-частицы;
$\overset{+}{\rightarrow} \Omega R$ [м]	точка в сеточном пространстве Ω на следующий (+) шаг во времени.

Богданов Александр Владимирович — доктор физико-математических наук, профессор, bogdanov@csa.ru

Дегтярев Александр Борисович — доктор технических наук, профессор, deg@csa.ru

Храмушин Василий Николаевич — кандидат технических наук, докторант, khram@mail.ru

ПРЕДИСЛОВИЕ

История становления CFD проходила в видении перспектив интенсивного роста вычислительных ресурсов и периодических оценок, сколько тера- или эксафлопов понадобится, чтобы решить все проблемы этой науки весьма архаичными математическими методами. На этом фоне резким диссонансом выглядели попытки российских учёных [Дородницын, 1997] вернуться к постановке вычислительной задачи на основе синтеза исходных физических принципов, подойти с новых позиций к проектированию вычислительных экспериментов и выходить на прямое моделирование течений. Это, однако, списывалось на нехватки вычислительной техники у таких исследователей, а все силы вновь и вновь бросались на создание морально устаревающих пакетов стандартных программ, с не всегда адекватными алгоритмами, и на создание далёких от инженерных видений инструкций по их использованию. То, что эти пакеты программ не всегда корректно описывали физическое явление и безвозвратно утрачивали связь с инженерными достижениями в архитектуре вычислительных систем, по инерции списывается на нехватку ресурсов, и потому современные книги по CFD смотрятся именно как описания соответствующих стандартных программ без должной взаимосвязи с их прикладным предназначением. Впрочем, глубокий анализ чисто физических проблем в создании математических моделей и проектировании прикладных алгоритмов в CFD был выполнен ещё в 1974 г. [Astarita, Marrucci, 1974].

Первые предостерегающие сигналы прозвучали в начале века, когда оказалось, что большинство стандартных программ, если и портируются на серверы с многоядерными процессорами, то, во всяком случае, очень плохо масштабируются. В настоящее время, когда все большие вычислительные комплексы гетерогенны, никаких идей по переносу на них стандартных CFD программ не просматривается. Причина этого понятна — задачи динамики элементарного объёма и её геометрические аспекты так сильно завязаны, что в процессе вычислений на каждом такте необходимо обмениваться таким количеством разнородной информации, которое сравнимо со всеми обрабатываемыми данными. Но ни один гибридный процессор не способен обеспечить этого. К настоящему времени стала очевидной недостаточность формального распараллеливания с помощью векторизации вычислительных потоков, что обусловлено необходимостью повсеместного контроля и гибридной перенастройки вычислительных процессов в зависимости от физического состояния моделируемых жидкостей. Развитие континуально-корпускулярных методов на основе российского «Метода крупных частиц» [Белоцерковский, Давыдов, 1982], приводит к вычислительным моделям тензорной математики с независимым контролем состояния каждой расчётной ячейки-частицы жидкости. Вычислительные алгоритмы и функциональная логика синтеза физических явлений и процессов для этой ячейки обеспечивается параллельно работающими арифметико-логическими ядрами, что в точности соответствует тенденциям развития вычислительной техники по запросам графической визуализации трёхмерных пространственных явлений и динамических процессов с ними. Именно использование тензорной алгебры для прямого моделирования физических явлений и процессов в составе обобщённой тензорной математики, позволяет эффективно синтезировать гидродинамические и геометрические аспекты вычислительного процесса в целом. Это давно поняли в теории поля. Такая программа для квантовой гравитации была намечена в работе [Храмушин, 2005], и блестяще реализована в статье [Wesseling, 2001]. Проблемы в CFD имеют ту же природу, поэтому естественно применять и здесь подход, основанный на тензорной математике, а построение алгоритмов для работы с тензорными числовыми объектами на современных гибридных системах реализуется естественным образом.

КЛЮЧЕВЫЕ ОПРЕДЕЛЕНИЯ

Алгоритмическая реализация прямого численного моделирования нестационарных процессов в гидромеханике с использованием явных схем имеет прекрасный исторический аналог в форме исчисления флюксий Исаака Ньютона, что в современных алгоритмах представляется числовыми объектами трёхмерного пространства с виртуальными контекстными связями для реализации законов гидроаэромеханики и физической теории поля на основе операций тензорной математики.

Исчисление флюксий* создаёт основу классической механики Ньютона для движения в векторном пространстве и скалярном времени [Эглит, 2008]. В кинематике первой флюксией становится скорость движения \vec{V} [м/с], как разностный дифференциал (*по Ньютону — момент*) в произведении с шагом во времени Δt или просто t [с]. Для подвижной точки a [м⁻²], замороженной в жидкость, кинематическое уравнение включает дифференциал (*момент*) скоростей движения и времени в единой абсолютной системе координат, так же как и внутри локального деформируемого базиса, связанного с конкретной крупной частицей жидкости — виртуальным числовым объектом**:

$${}^+A = R + \vec{V} \cdot t + a \cdot (\mathbf{r} + \mathbf{v} \cdot t) + \omega \cdot (\mathbf{r} + \mathbf{v} \cdot t)t \text{ [М]}, \quad (1)$$

где ω — скорость смещения точки внутри частицы жидкости; ${}^+A$ — новое положение контрольной точки в абсолютной системе координат; R — местоположение локального базиса в абсолютной системе координат; \mathbf{v} — тензор вращения и деформации базисных осей тензора формы \mathbf{r} .

На сопряжённом этапе вычислительного эксперимента используются обновлённые пространственные поля скоростей и деформаций, необходимые для контроля динамического состояния и переопределения реологических характеристик жидкости в форме скалярных, векторных и тензорных неразрывных физических полей следующего расчётного интервала во времени. В современных интегральных вычислительных моделях аналогичные гидродинамические особенности ассоциируются с пространственными распределениями поляризованных дипольных ядер, с производными от них вихресточниками на поверхностях разрыва сплошности или устойчивости течений реальной вязкой, упругой и сжимаемой*** жидкости.

* Флюенты (*fluens, fluentis* — текущий) — функции x, y, z по аргументу времени t ; Флюксии (*fluxio* — истечение) — $\dot{x}, \dot{y}, \dot{z}$ производные x, y, z по времени t .

$${}^+A_x = {}^oA_x + (V_x + v_{xx}a^x + v_{xy}a^y + v_{xz}a^z)t;$$

** В скалярной записи: ${}^+A_y = {}^oA_y + (V_y + v_{yx}a^x + v_{yy}a^y + v_{yz}a^z)t;$

$${}^+A_z = {}^oA_z + (V_z + v_{zx}a^x + v_{zy}a^y + v_{zz}a^z)t.$$

*** В тензорной математике определяются реологические характеристики для вязкой и упругой жидкости, зависящие от скорости и величины локальной деформации, соответственно.

ГЕОМЕТРИЧЕСКОЕ ОБОБЩЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ ОБЪЕКТОВ И СВЯЗАННЫХ АЛГОРИТМИЧЕСКИХ ОПЕРАЦИЙ

Геометрическое построение пространственной задачи включает скалярные, векторные и тензорные числовые объекты с определённым выбором алгоритмических процедур и арифметико-логических операций, в размерной физической форме и с привязкой числовых объектов к интерполяционным базисам в тензорном* сеточном пространстве.

Элементарные числовые объекты образуются симплексами некопланарных базисных векторов (см. рис. 1), используемых в качестве линейных интерполяционных операторов — разностных дифференциалов в трёхмерном евклидовом пространстве \vec{r}_i и служат для построения неразрывных физических полей в окрестности смежных сеточных узлов ${}^T\Omega R$ и центров масс ω для сопряжённых с ними крупных деформируемых и подвижных частиц сплошной среды. Произведение векторных и тензорных величин выполняются со свёрткой — суммированием немых индексов в одночленном произведении** $a = a \cdot r$ или $a^j = a_i r^{ij}$ — при переходе в локальный базис, и обратно — $a = a \cdot r$ или $a_k = a^j r_{jk}$ — при возврате к абсолютным координатам (тензор локального базиса — справа от вектора).

- *Левый верхний индекс* отмечает текущее время, которое может обозначаться прописной буквой ${}^T\Omega$ в абсолютном измерении, или расчётным шагом во времени tR , или значками ${}^+\omega$ и ${}^-\omega$ — для ссылок на следующий или предыдущий интервал во времени;
- *Левый нижний индекс* — местоположение в сеточном пространстве $\{X, Y, Z\}r$, или ссылки на смежные узлы $\{+\}r$ или центры масс частиц жидкости $\{-\}r$ на сопряжённых этапах вычислительного эксперимента.

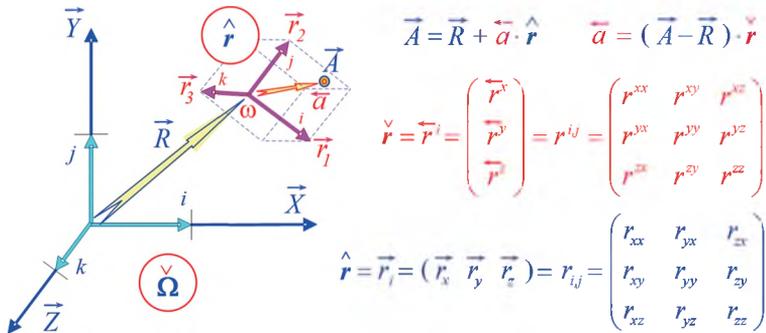


Рис. 1. Геометрия глобального сеточного пространства $\check{\Omega}$ и множества локальных базисов $\hat{r}(\omega)$; i, j, k — единичные орты в связанной системе отсчёта

* Тензорное или евклидово пространство изнутри опирается на ортогональную декартову систему координат с ортами в масштабе исходных базисных векторов.

** Тензорное произведение служит для смены сопряжённых координатных базисов, или для изменения физических свойств пространственных числовых объектов.

Правые индексы связывают векторные и тензорные компоненты в абсолютном и локальных базисах, что служит строгому определению динамики и деформации числовых ячеек — частиц сплошной среды.

- Нижние правые индексы, тензорный «домик» и стрелочка вправо показывают принадлежность единой абсолютной системе координат (см. рис. 1).

Тензор $\overset{\wedge}{\mathbf{r}}$ [м³] собирается по столбцам из базисных векторов \vec{r}_i в матрице геометрических преобразований типа $\vec{a} = \overset{\leftarrow}{a} \cdot \overset{\wedge}{\mathbf{r}}$ [м].

- Верхние правые индексы — проекции внутри тривектора подвижной и деформируемой сеточной ячейки, где отображение ортов абсолютных координат ложится векторами-строками в матрице обратных координатных преобразований $\overset{\vee}{\mathbf{r}} = \overset{\leftarrow}{\mathbf{r}}^j = \overset{\wedge}{\mathbf{r}}^{jk} = \overset{\leftarrow}{\mathbf{r}}^{-1}$ [м⁻³], отмечаемых тензорной «галочкой» и векторной стрелочкой влево $\overset{\vee}{a} = \overset{\leftarrow}{a} \cdot \overset{\wedge}{\mathbf{r}}$ [м⁻²].
- Прописными буквами отмечаются большие числовые величины, измеряемые в масштабной шкале глобального пространства Ω и единого абсолютного времени T ;
- Строчными — существенно малые величины или конечные разности, соизмеримые с физическими размерениями локальных базисов элементарных частиц сплошной среды ω , и в интервале текущего шага времени t .

Абсолютное время T может содержать юлианскую дату и время от начала суток* ${}^k T = T + kt$, так же как и абсолютные отсчёты в пространстве** $A = R + \vec{a} \cdot \overset{\wedge}{\mathbf{r}}$ [м] — географические и другие обобщённые координаты. В сбалансированных численных схемах исключается необходимость вовлечения абсолютных отсчётов в пространстве ${}^T \vec{R}$ и времени T , так как на всех этапах вычислений достаточно использования числовых величин в узлах и центрах масс сопряжённых сеточных ячеек — виртуальных корпускул $\overset{\wedge}{\mathbf{r}} = \overset{\circ}{R} - \overset{\circ}{R}$ [м³].

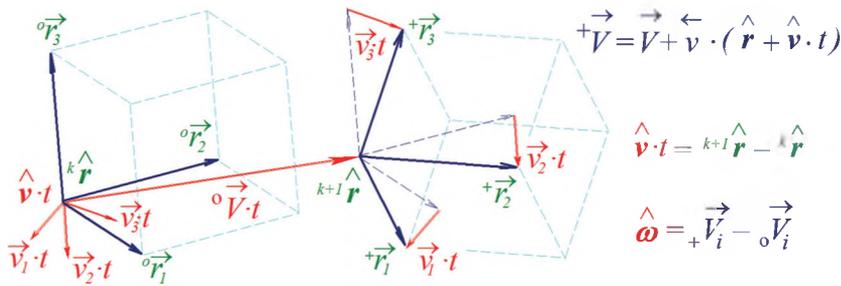


Рис. 2. Тензор изменения скоростей $\overset{\wedge}{\omega}$ [м³/с] образуется деформационными движениями базисных векторов расчетной ячейки в пространстве, на фоне их смещения в пространстве $\overset{\vee}{\mathbf{v}} \cdot t$ [м³] за интервал времени t [с] (Тензоры $\overset{\wedge}{\omega}$ и $\overset{\vee}{\mathbf{v}}$ могут быть вырожденными, а их обращение — некорректно.)

* Истинное время задаётся числовой структурой **Event** с юлианской датой: **D** (от 4713 г. до н.э.), и локальным временем в часах от начала суток **T**.

** В программной среде точки в глобальных координатах: **Point** — разделены со свободными векторами в локальных базисах: **Vector**, что унифицирует вычислительные операции с тензорными числовыми объектами: **Tensor** и **Basis**.

Кинематика внутренних потоков задаётся разностным тензором скорости (рис. 2) по смещаемым во времени базисным векторам формы частицы жидкости:

$$\overset{\wedge}{\mathbf{v}} \cdot \vec{t} = v_i \cdot \vec{t} = \overset{\Delta}{\Delta} \vec{r}_i = \overset{t}{+} \overset{r}{r}_i - \overset{o}{o} \overset{r}{r}_i \quad [\text{м}^3], \quad (2)$$

где тензор $\overset{\wedge}{\mathbf{v}}$ [м³/с] задаёт текущие скорости смещения базисных ортов частицы жидкости в локальном масштабе (*строчные буквы*), измеренные в проекции глобальной системы координат (*индексы снизу*). Преобразованием системы отчёта тензора скоростей* к локальному базису крупной частицы жидкости (*геометрическая нормализация*) $\overset{<}{\mathbf{v}} = \overset{\wedge}{\mathbf{v}} \cdot \overset{\wedge}{\mathbf{r}}$ [1/с] получается независимый тензор конвективных скоростей, характеризующий локальное движение жидкости. Тензор $\overset{<}{\mathbf{v}}$ [1/с] содержит расширенный набор кинематических элементов системы дифференциальных уравнений с перекрёстными производными компонентами движения деформируемой частицы жидкости:

$$\overset{<}{\mathbf{v}} = \overset{\wedge}{\mathbf{v}} \cdot \overset{\wedge}{\mathbf{r}} = \begin{pmatrix} v_{xx} r^{xx} + v_{xy} r^{yx} + v_{xz} r^{zx} & v_{xx} r^{xy} + v_{xy} r^{yy} + v_{xz} r^{zy} & v_{xx} r^{xz} + v_{xy} r^{yz} + v_{xz} r^{zz} \\ v_{yx} r^{xx} + v_{yy} r^{yx} + v_{yz} r^{zx} & v_{yx} r^{xy} + v_{yy} r^{yy} + v_{yz} r^{zy} & v_{yx} r^{xz} + v_{yy} r^{yz} + v_{yz} r^{zz} \\ v_{yz} r^{xx} + v_{zy} r^{yx} + v_{zz} r^{zx} & v_{yz} r^{xy} + v_{zy} r^{yy} + v_{zz} r^{zy} & v_{yz} r^{xz} + v_{zy} r^{yz} + v_{zz} r^{zz} \end{pmatrix}$$

Иначе такое произведение можно представить в форме полного дифференцирования $\overset{<}{\mathbf{v}} = \overset{\wedge}{\mathbf{v}} / \overset{\wedge}{\mathbf{r}} = \overset{\Delta}{\Delta} \vec{v}_i / \overset{\Delta}{\Delta} \vec{r}_i$, исполняемого без искусственных исключений «малых» или конвективных элементов в приближениях субстанциональных производных, и при корректной физической интерпретации реологических характеристик жидкости и условий существования течений, с возможностью асимптотических или эмпирических уточнений численных моделей.

МАТЕМАТИЧЕСКАЯ ФОРМАЛИЗАЦИЯ КONTИНУАЛЬНО-КОРПУСКУЛЯРНОГО МОДЕЛИРОВАНИЯ В ГИДРОМЕХАНИКЕ

Тензорная математика** ориентирована на создание прямых вычислительных экспериментов в решении прикладных задач гидромеханики. Континуально-корпускулярный подход строится по численным схемам первого порядка с последовательным разностным интегрированием законов движения на сопряжённых этапах по скалярному аргументу — времени. Разделение этапов вычислений по совокупным физическим процессам даёт возможность сквозного контроля и гибридного перестроения математических зависимостей по текущим оценкам состояния моделируемой сплошной среды с учётом интенсивности физического взаимодействия смежных корпускул, как виртуальных числовых объектов. Каноническое представление законов гидромеханики позволяет строго и однозначно связывать числовые объекты с арифметико-логическими операциями и сложными геометрическими алгоритмами, в том числе с использованием быстрых интерполяций для нерегуляризованных сеточных пространств.

* Запрет повышения ранга в операциях произведения позволяет использовать автоматические перестановки сомножителей в геометрических преобразованиях $\overset{\wedge}{\mathbf{a}} \leftarrow \overset{\wedge}{\mathbf{a}} \overset{\rightarrow}{\mathbf{v}} \overset{\wedge}{\mathbf{v}} \rightarrow \overset{<}{\mathbf{a}}$
 $\overset{\wedge}{\mathbf{v}} \cdot \overset{\wedge}{\mathbf{a}} = \overset{\wedge}{\mathbf{v}} \cdot (\overset{\wedge}{\mathbf{a}} \cdot \overset{\wedge}{\mathbf{r}}) = \overset{\wedge}{\mathbf{v}} \cdot \overset{\wedge}{\mathbf{r}} \cdot \overset{\wedge}{\mathbf{a}} = \overset{\wedge}{\mathbf{v}} \cdot \overset{\wedge}{\mathbf{a}} \quad [\text{м/с}].$

** Математика — наука о количественных отношениях и пространственных формах действительного мира.

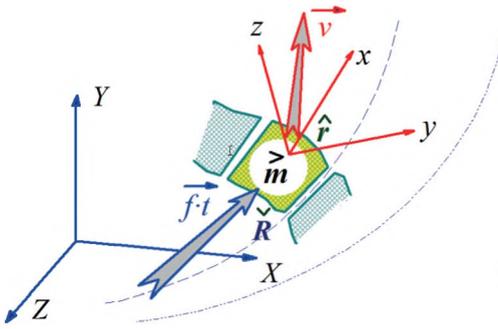


Рис. 3. Разделение этапов вычислений по физическим процессам включает тензорную массу \hat{m} в качестве геометрического оператора, корректирующего движение частиц жидкости по живым силам и инерции на сопряжённых этапах моделирования во времени

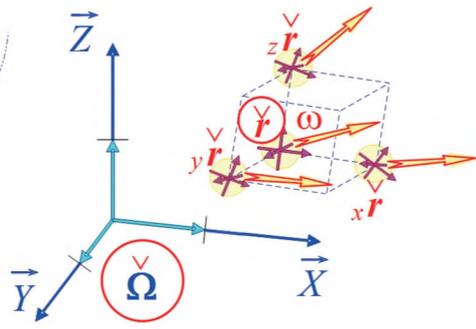


Рис. 4. На корпускулярном этапе вычислительного эксперимента выполняется оценка текущего реологического состояния сплошной среды в исходных узлах расчётной сетки

Последовательное пространственное интегрирование первого порядка задействует сеточные и корпускулярные подходы, формализуемые с помощью тензора инерционной массы $\hat{m} = m_k^i$ [кг] в качестве числового объекта для фиксации предыстории — инерции в движении и локальной деформации моделируемых частиц жидкости, что образует алгоритмическую последовательность сбалансированного прогноза и коррекции движения по криволинейным траекториям частиц жидкости с динамически изменяющейся внутренней энергией.

Внешняя массовая сила в абсолютной системе координат \vec{f} [кг·м/с²] в векторной трактовке второго закона Ньютона представляется производной приращения количества движения (строчная буква в обозначении скорости) $\vec{v} \cdot \hat{m} / t$ во времени:

$$\vec{f} \cdot t = \vec{v} \cdot \hat{m} \text{ [кг·м/с]}.$$

На континуальном этапе вычислительного эксперимента (рис. 3, 4) [Храмушин, 2005] для каждого пространственного сеточного узла ΩR [м] уточняются векторные поля абсолютных \vec{V} [м/с] и локальных скоростей \vec{v} [м/с], находящихся в ближайших окрестностях расчётной ячейки \hat{a} [м⁻²] с учётом деформации и изменения внутренней кинетической энергии $\hat{\omega}$ [с⁻¹] под действием внешних массовых сил \vec{f} [кг·м/с²] и поверхностных напряжений \hat{f} [Н·м²], действующих в условиях активного влияния внутренней энергии и инерции $\hat{m} = \rho \cdot \hat{r}$ [кг] конкретной частицы \hat{r} [м³] в произведении с тензором плотности ρ [кг/м³], формализующим предысторию действия объёмной силы инерции и поверхностного комплекса живых сил — собственной внутренней энергии жидкой корпускулы:

$$\vec{+}A = \vec{R} + \vec{V} \cdot t + \frac{\vec{m} \cdot \vec{f} \cdot t^2}{2} + \left| \frac{\overset{\wedge}{\mathbf{r}} + \overset{\wedge}{\mathbf{v}} \cdot t + \overset{\wedge}{\mathbf{m}} \cdot \vec{f} \cdot t^2}{2} \right| \cdot \overset{\leftarrow}{\mathbf{a}} \quad [M], \quad (3)$$

где $\vec{+}A$ [M] — местоположение точки в новом расчётном шаге во времени; \vec{R} [M] — исходное местоположение центра отсчёта для числовой ячейки; \vec{V} [м/с] — скорость свободного движения частицы жидкости; \vec{f} [Н] — внешняя объёмная сила; $\overset{\wedge}{\mathbf{f}}$ [Н·м²; кг·м³/с²] — тензор поверхностных напряжений, предопределённый из граничных условий. В случае использования специальных пространственно разнесённых сеточных шаблонов, интерполяционные поправки на локальное местоположение центра масс $\overset{\leftarrow}{\mathbf{a}}$ [м⁻²] могут стать пренебрежимо малыми.

Сопряжённый корпускулярный этап вычислений получает обновлённое кинематическое поле тривекторов локальных скоростей $\overset{\wedge}{\mathbf{v}}$ [м³/с] с уточнением деформационных смещений относительно подвижных корпускул $\overset{\wedge}{\varepsilon}$ [м³] для обратного интерполирования — вторичного разностного интегрирования (рис. 4) с фиксацией напряжений $\overset{\leftarrow}{\sigma}$ [кг·м²/с²] в ближайших окрестностях виртуальных числовых объектов:

$$\overset{\leftarrow}{\sigma} = (\overset{\wedge}{\varepsilon} \cdot \overset{\wedge}{\kappa} + \overset{\wedge}{\mathbf{v}} \cdot \overset{\wedge}{\eta}) \times \lambda \quad [H \cdot m], \quad (4)$$

где $\overset{\wedge}{\kappa}$ [Н/м³] — тензорный модуль упругой деформации; $\overset{\wedge}{\eta}$ [Н·с/м³] — тензор коэффициентов вязкости; λ [м] — масштабный коэффициент, характеризующий дистанцию между взаимодействующими виртуальными энергетическими объектами — крупными частицами жидкости.

Заметим, что тензоры напряжений и локальных скоростей могут быть вырожденными. Без возможности их обращения $\overset{\wedge}{\mathbf{v}} \cdot \overset{\leftarrow}{\sigma}$ выполняется только в операциях произведения с тензорами формы $\overset{\wedge}{\mathbf{v}} \cdot \overset{\leftarrow}{\sigma} = \overset{\leftarrow}{\mathbf{v}} \quad [c^{-1}]$, либо путём алгоритмического синтеза по интерполяционным оценкам состояния окружающей жидкости.

По аналогии с работой [Эглит, 2008], внутрь подвижных частиц жидкости вовлекается скалярная функция распределения температур, растекающихся между свободными частицами по отрицательным градиентам; или связанное скоростью звука изменение плотности упругого газа в функции от дивергенции скорости и величины давления, в том числе с уточнениями по закону Бернулли и с учётом внутренней тепловой энергии, что в совокупности синтезирует расширенное решение прикладной задачи аэрогидромеханики.

В качестве основного цикла в вычислительном эксперименте принимается последовательность алгоритмов для согласования кинематических и реологических характеристик физического поля в явных численных схемах, сводимых к режиму последовательного установления нестационарных вычислительных процессов, происходящих по математическим моделям для вязких, упругих и сжимаемых жидкостей.

ТЕНЗОРНОЕ ПРЕДСТАВЛЕНИЕ ОСНОВНЫХ ЗАКОНОВ ГИДРОМЕХАНИКИ

Формальное построение физических объектов и операций в тензорной математике [Храмушин, 2005] приводит к строгим определениям для свособразной «модели мира» вычислительной гидромеханики:

- континуально-корпускулярная вычислительная модель метода «Крупных частиц» [Белоцерковский, Давыдов, 1982] в тензорной записи сводится к двойной линейной разностной интерполяции физических полей (*место интегрирования уравнений движения второго порядка*);
- движение и взаимодействие крупных частиц жидкости строится в операциях произведения, что более соответствует физике пространственных процессов (*нет ограничений по малости дифференциальных аппроксимаций*);
- использование явных численных схем и дискретных числовых полей служит повышению эффективности прямых вычислительных экспериментов, и не исключают возможности контроля корректности и, по необходимости, задействования гибридных схем для достижения адекватных инженерных результатов прямого численного моделирования.

Закон движения для свободной частицы сплошной среды (см. рис. 3):

$$\vec{f} = \vec{w} \cdot \vec{m} = \vec{w} \cdot \rho \cdot \vec{r} \quad [\text{Н}, \text{кг} \cdot \text{м} / \text{с}^2]; \quad (5)$$

Тензор вязких напряжений в потоке локальных скоростей (см. рис. 4):

$$\vec{f}_H = \vec{v}_H \cdot \vec{\eta} / \lambda = \vec{v}_H \cdot \vec{\eta} / \lambda \quad [\text{Н} / \text{м}]. \quad (6)$$

Тензор упругих напряжений при деформации корпускулы*:

$$\vec{f}_T = \vec{v}_T \cdot \vec{\varkappa} / \lambda = \vec{v}_T \cdot \vec{\varkappa} / \lambda \quad [\text{Н} / \text{м}], \quad (7)$$

где тензор локальных скоростей $\vec{v} = \overset{\wedge}{+} \vec{V}_i - \overset{\circ}{-} \vec{V}_i$ [м³/с] (см. рис. 2); реологические тензоры: $\vec{m} = \overset{\vee}{\rho} \cdot \vec{r}$ [кг] — тензор массы–инерции; \vec{r} [м³] — тензор формы; ρ [кг/м³] — «условная плотность» сохранения предыстории девиаций — внутренних «живых сил» подвижных корпускул; $\vec{\eta}$ [кг/с/м²] — тензоры динамической вязкости и жёсткости $\vec{\varkappa}$ [кг/м²] реальной жидкости, λ [м] — дистанция ближнего взаимодействия смежных частиц.

Вычислительная модель содержит реологические свойства жидкости: вязкость и упругость, при этом соотношение интенсивностей получающихся напряжений определяет возникновение критических режимов течения с образованием струй, вихревых слоёв и кавитационных разрывов. Под действием тензора внутренних напряжений, частица жидкости получает приращение скорости внутреннего (замкнутого) движения, традиционно представляемого с помощью трёх динамических состояний:

$$\vec{f} = \vec{v}_0 \cdot \vec{\varepsilon} \cdot \vec{t} + \vec{v}_H \cdot \vec{\mu} / \lambda + \vec{v}_T \cdot \vec{\varkappa} / \lambda = \vec{f}_0 + \vec{f}_H + \vec{f}_T \quad [\text{Н} / \text{м}], \quad (8)$$

* Реологические параметры жидкости в тензорной математике не сохраняют форму традиционных выражений, и для согласования физических размерностей введена характерная скалярная дистанция λ [м] в записях законов вязкости и упругости.

где тензоры $\overset{<}{f}_0$ — давление; $\overset{<}{\varepsilon}$ — коэффициент сжатия. Получающийся характеристический полином может использоваться в оценках состояния жидкости, где реологические параметры проявляются в виде главных инвариантов тензора конвективных скоростей:

- $\overset{<}{v}_0 : \text{I} \neq 0$ — сжимаемость \rightarrow кавитационный разрыв плотности;
- $\overset{<}{v}_H : \text{II} \neq 0$ — поворот \rightarrow образование свободной струи или турбулентного вихря;
- $\overset{<}{v}_T : \text{III} \neq 0$ — деформация, чистая — если другие инварианты тензоров равны нулю.

Так, если расчётная ячейка с присоединённым вихрем обращает в нуль детерминант внутреннего поля конвективных скоростей в смежной точке — в точке центра масс свободной сопряжённой частицы жидкости, что предопределяет зарождение свободного турбулентного вихря внутри крупной частицы жидкости. Такой гибридный алгоритм можно использовать за пределами аппроксимационного разрешения на относительно грубых сетках, либо для уточнения решения, когда сеточная область в зоне вихреобразования должна сгущаться.

ОБ ОСОБЕННОСТЯХ ТЕНЗОРНОГО ОБОБЩЕНИЯ ПРИКЛАДНОЙ ЗАДАЧИ ГИДРОМЕХАНИКИ

На этапах вычислительного эксперимента поле скорости $\overset{\wedge}{\omega}$ [м³/с] формализуется с использованием потенциально вырожденных тензоров, что вполне удовлетворяет традиционному аддитивному построению модели течения $\overset{+}{r} = \hat{o} r + \hat{v} \cdot t$ [м³] в абсолютном времени t [с]. С целью сопоставления логики вычислительных операций с традиционными понятиями механики сплошных сред допустим справедливость определений для тензора — аффинора в глобальной системе координат, и разложение Коши-Гельмгольца [Готман, 2007] для пространственных смещений: расширение (*дивергенция*); закрутка (*ротор*) и деформация (*сдвиг*).

Выделением диагонального тензора \hat{v}_o [м³/с], такого, что след девиаторного остатка $\text{tr } \hat{v}^*$ будет равен нулю [Филин, 1975], получим тензор шарового сжатия

$$\left. \begin{array}{l}
 \text{[Иллюстрация тензора сжатия: звезда с 6 лучами, сходящимися к центру, с чередующимися цветами (синий, желтый, синий, желтый, синий, желтый)]} \\
 \hat{v} = \hat{v}_o + \hat{v}^* \quad (\text{tr } \hat{v}^* = 0) \text{ [м}^3\text{/с]}; \\
 \overset{<}{\sigma} = \hat{v}_o \cdot \overset{\vee}{\varepsilon} \cdot t \text{ [кг/с}^2\text{]},
 \end{array} \right\} \quad (9)$$

где детерминант тензора $\overset{<}{\sigma}$ [кг/с² = Н/м] определяет давление, $\overset{\vee}{\varepsilon}$ [Н/м²] — коэффициент объёмного сжатия — модуль продольной упругости.

Кососимметричная суперпозиция с компонентами девиаторного тензора \hat{v}^* выделяет начальный импульс к вращению частиц жидкости с возникновением напряжения $\overset{<}{f}_H$ в ньютоновском определении механики вязкой жидкости (6):

$$\text{[Иллюстрация тензора напряжения: три вращающихся стрелки в центре, сходящиеся к центру, с чередующимися цветами (синий, желтый, синий)]} \quad \overset{<}{f}_H = \hat{v}_H \cdot \overset{\vee}{\eta} / \lambda = (\hat{v}^* - \hat{v}^{*T}) \cdot \overset{\vee}{\eta} / (2\lambda) \text{ [кг/с}^2\text{]} \quad (10)$$

с матрицей коэффициентов вязкости $\overset{\vee}{\eta}$ [Н·с/м³] в произведении с тензором скоростей локальных деформаций \hat{v}_H [м³/с].

Симметризация девиаторного тензора $\overset{\wedge}{\nu}^*$ [м³/с] связывается со сдвиговыми напряжениями $\overset{<}{f}_T$ [Н/м²], возникающими внутри крупной частицы сплошной среды по закону Гука (7):

$$\overset{<}{f}_T = \overset{\wedge}{\nu}_T \cdot \overset{<}{\kappa} \cdot t / \lambda = \left| \overset{\wedge}{\nu}^* + \overset{\wedge}{\nu}^{*T} \right| \cdot \overset{<}{\kappa} \cdot t / (2\lambda) \quad [\text{кг/с}^2], \quad (11)$$

с модулем упругости $\overset{<}{\kappa}$ [Н/м³] и симметричным тензором относительных локальных смещений в окрестности крупной частицы жидкости $\overset{\wedge}{\nu}_T \cdot t$.

Реологические характеристики состояния жидкости — $\overset{<}{\varepsilon}$ [Н/м²], $\overset{<}{\kappa}$ [Н/м³] и $\overset{<}{\eta}$ [Н·с/м³] — могут также задаваться скалярными коэффициентами, аналогами шаровых тензоров в традиционных канонических зависимостях между напряжениями, локальными деформациями и вязкостными движениями сплошной среды. В тензорном обобщении неявно представляется анизотропия практического связывания перекрёстных компонент для взаимовлияния внутренних напряжений в реологических характеристиках жидкости, представляемых традиционных в скалярных описаниях с помощью безразмерных коэффициентов Пуассона μ .

Полный тензор внутренних напряжений однозначно связывает геометрические смещения, скорости деформаций и приращения скоростей с поверхностными напряжениями:

$$\overset{<}{\Phi} = \overset{<}{\sigma} + \overset{<}{f}_T + \overset{<}{f}_H = \left| \overset{\wedge}{\nu}_0 \cdot \overset{<}{\varepsilon} + \overset{\wedge}{\nu}_T \cdot \overset{<}{\kappa} / \lambda \right| \cdot t + \overset{\wedge}{\nu}_H \cdot \overset{<}{\eta} / \lambda \quad [\text{Н/м}^2]. \quad (12)$$

Для приведения закона движения с инерционной массой $\overset{<}{f} \cdot t = \overset{<}{m} \cdot \overset{<}{v}$ [кг/с] к новым зависимостям для поверхностных напряжений $\overset{<}{\Phi}$ [кг/м·с²] необходим сомножитель с размерностью пространственного вектора λ [м], который по физическому смыслу соответствует либо вектору направления движения для теоремы Жуковского о свободном вихредиполе, либо расстоянию между взаимодействующими частицами жидкости.

О ПОСТРОЕНИИ ТЕНЗОРНЫХ ЧИСЛОВЫХ ОБЪЕКТОВ И МОДЕЛИРУЮЩИХ АЛГОРИТМОВ

Геометрическое построение нерегуляризованных сеточных пространств [Wesseling, 2001] в явных численных схемах первого порядка для последовательного интегрирования в сопряжённых этапах континуально-корпускулярного вычислительного эксперимента при моделировании нестационарных процессов в гидромеханике представляется последовательным отображением архитектуры цифрового вычислительного комплекса:

- континуально-корпускулярное разделение этапов вычислительного эксперимента в гидромеханике строится в неразрывном пространстве физических величин, изначально определяемых в единой абсолютной системе координат с использованием скалярных, векторных и тензорных характеристик состояния и законов трансформации подвижной и деформируемой сплошной среды;
- в операциях произведения с участием внешних (массовых) и внутренних (поверхностных) напряжений за расчётный интервал времени перестраи-

вается внутреннее состояние каждой расчётной ячейки по образу независимой крупной частицы жидкости, аккумулирующей внутреннюю энергию и влияющей на кинематические реакции в интерполяционном переоброении обновляемых физических полей в исходных расчётных узлах;

- при весьма сложных физических законах гидромеханики включается дополнительный третий этап вычислений для контроля и корректировки внутреннего состояния частиц жидкости с согласованием законов сохранения массы и количества движения свободных частиц жидкости с позиций центровки вычислений на двух главных этапах вычислительного эксперимента по скалярному аргументу — времени.

Вычислительное пространство строится на неподвижных нерегуляризованных узлах — **Point**, в индексируемом множестве сеточных ячеек для трёхмерной интерполяции с опорой на евклидовы базисы — **Base**, внутри и в окрестности которых строится математика согласования физических законов с использованием свободных векторов — **Vector**.

Объектно-ориентированное программирование на алгоритмических языках C++ или D позволяет перенести на этап компиляции исходных текстов программ основной контроль корректности математических операций, если они применяются к однозначно определенным по физическим характеристикам числовым объектам, в том числе содержащим предварительно рассчитанные обратные матрицы, детерминанты или собственные значения и векторы, используемые в формировании конкретных математических моделей.

Основные вычислительные объекты (Программа построения числовых объектов и функций трёхмерной тензорной математики для вычислительных экспериментов в гидромеханике (Tensor). СПбГУ, Роспатент № 2013619727)] определяются из условия быстрого и независимого исполнения вычислительных операций и, по необходимости, с возможностью быстрой перестройки вычислений для применения гибридных алгоритмов в зависимости от текущих условий существования физического явления в локальной подобласти:

```
typedef double Real; // скалярная величина в глобальном пространстве и времени
typedef double real; // локальные или разностные отсчеты в пространстве
                    // и времени
struct Tensor;      // тензорный объект без контекстных связей для быстрого
                    // счета
struct Base;        // координаты местоположения и связный базис евклидовой
                    // СК
struct Cell;        // числовая ячейка со смежными контекстными связями
struct Point;       // точка дальних мировых отсчетов в масштабе абсолютной
                    // СК
struct Vector;      // свободный разностный вектор в локальном масштабе
struct Space;       // пространство узловых элементов сеточной области в целом
struct Volume;     // множество свободных — подвижных и деформируемых
                    // ячеек
```

Так, в случае автоматического преобразования сложных величин **Vector** или **Tensor** к локальной величине **real**, будут получаться длина вектора и детерминант матрицы тензора, всякие другие преобразования компилятор будет помечать ошибочными. Разность величин типа **Point** может дать только величину **Vector**, в то время как сложение объектов типа **Vector** с величинами **Vector** или **Point** приведёт тому же типу данных и никакому другому.

Современные компьютеры обладают практически неограниченными объёмами оперативной памяти и существенно малой скоростью арифметико-логических операций при использовании больших вычислительных массивов. Для технического обеспечения сквозного распараллеливания операций основные массивы данных **Space** и **Volume** должны сохраняться для текущего, предыдущего и последующего срезов во времени, что позволит синхронизировать параллельные вычисления в случае неполного разделения этапов эксперимента по независимым физическим процессам, или при необходимости создания каких-либо циклов для активного или релаксационного согласования параметров вычислительной среды в сложных и плохо обусловленных математических моделях.

В трёхмерной тензорной математике со всеми вычислительными объектами должны связываться логические предикаты (*token-признаки*) для установки типа числовых структур, их принадлежности к абсолютному или локальному базису, и которые повсеместно используются для автоматических преобразований в соответствии с исходными установками в уравнениях аэрогидромеханики. Как минимум, можно сформировать базовые значения *s*-предикатов, которые будут включаться во все числовые объекты в виде особой тройки двоичных бит:

- “000” — \vec{T} — скалярная величина в абсолютном базисе или инвариант тензора;
- “001” — $\overset{\leftarrow}{A}$ — вектор в глобальной системе координат (СК);
- “010” — a — вектор в локальном базисе крупной частицы;
- “011” — \hat{t} — размерная скалярная величина в локальной системе отсчета;
- “100” — $\overset{>}{r}$ — тензор формы крупной частицы жидкости в глобальной СК;
- “101” — $\overset{<}{v}$ — смешанный базис проекций векторов на локальную СК;
- “110” — $\overset{\vee}{m}$ — базис локальных векторов в проекциях глобальной СК;
- “111” — ρ — определение тензорной величины в локальном базисе.

Непрерывный контроль битовых масок-признаков особенно важен в случае глубокого распараллеливания вычислительных операций на сложных гибридных системах, разделяющих моделирование на большие пространственные блоки, между которыми должна осуществляться автоматическая конвертация числовых структур для согласования решений на смежных границах или в областях с перелестом нерегулярных сеточных аппроксимаций, и где возможны непредсказуемые перестроения структуры физических полей аэрогидромеханики.

ЗАКЛЮЧЕНИЕ

Из необходимости адаптивного управления вычислительным экспериментом со сквозным контролем текущего состояния всех числовых объектов и операций устанавливаются функциональные или контекстные требования к специализированной среде программирования для проектирования, построения и реализации прямых вычислительных экспериментов в континуально-корпускулярных физических средах:

- элементарные пространственно-временные объекты и базовые физические явления записываются тензорными величинами в размерном виде, что допускает визуальный контроль и применение гибридных схем с инженерными эмпириками и математическими асимптотиками;

- вычислительные операции инвариантно определяются в проекциях глобальной системы координат и ассоциируются обратными аппроксимациям в локальных базисах, геометрически связанными произведениями числовых объектов — крупных частиц, конечных объёмов жидкости.

Объектно-ориентированная среда прикладного программирования связывает числовые объекты арифметико-логическими операциями:

- 1) логические или эмпирические алгоритмы формализуют законы прикладной гидромеханики, способы формирования и методы анализа текущего состояния сплошной среды в условиях интенсивной трансформации числовых объектов;
- 2) сложение применяется к числовым объектам в едином геометрическом базисе при строгом соответствии физических размерностей, в рамках предопределённых законов гидромеханики (п. 1);
- 3) произведение выполняется с векторами и тензорами в дуальных координатных базисах, что исключает повышение ранга в одночленных произведениях, если это не покрывается синтезом алгоритмов (п. 1) или анализом числовых объектов (п. 2).

Алгоритмические последовательности управляют числовыми объектами, формализующими реологические свойства и законы движения в континуально-корпускулярном представлении механики сплошных сред:

- скалярные или инвариантные величины, например время t , участвуют в операциях произведения или представляются производными от любых числовых объектов;
- векторные величины участвуют в операциях сложения только с соизмеримыми векторами или в операциях произведения с тензорами для изменения координатных базисов;
- тензорные величины синтезируют числовые объекты или характеризуют состояние крупных частиц жидкости, представляя геометрические деформации в качестве физических явлений и процессов гидромеханики в функции скалярного времени.

С использованием тензорных объектов формулируются основные законы гидромеханики, а их виртуальное конструирование и текущий анализ состояния образуют этапы прямого численного моделирования.

ЛИТЕРАТУРА

- [Белоцерковский, Давыдов, 1982] *Белоцерковский О. М., Давыдов Ю. М.* Метод крупных частиц в газовой динамике. Вычислительный эксперимент. М.: Наука, 1982. 370 с.
- [Готман, 2007] *Готман А. Ш.* Тензорное исчисление. Новосибирск: Новосибирская государственная академия водного транспорта, 2007. 129 с.
- [Дегтярев, Храмушин, 2014] *Дегтярев А. Б., Храмушин В. Н.* Проектирование и построение вычислительных экспериментов в гидромеханике с использованием явных численных схем и алгоритмов тензорной математики // Математическое моделирование. 2014. Т. 26. № 11. С. 4–17.
- [Дородницын, 1997] *Дородницын А. А.* Избранные научные труды. Т. 1, 2. М.: ВЦ РАН, 1997. Т. 1. 395 с.; Т. 2. 351 с.

- [Мак-Коннел, 1963] *Мак-Коннел А. Дж.* Введение в тензорный анализ. М.: Наука, 1963. 411 с.
- [Филин, 1975] *Филин А. Н.* Прикладная механика твёрдого деформируемого тела. Т. 1. М.: Наука, 1975. 832 с.
- [Храмушин, 2005] *Храмушин В. Н.* Трёхмерная тензорная математика вычислительных экспериментов в гидромеханике. Владивосток: ДВО РАН, 2005. 212 с.
- [Эглит, 2008] *Эглит М. Э.* Лекции по основам механики сплошных сред. М.: Изд-во Моск. ун-та, 2008. 318 с.
- [Astarita, Marrucci, 1974] *Astarita G., Marrucci G.* Principles of non-Newtonian fluid mechanics. McGRAW-HILL Book Company (UK), 1974. 289 p.
- [Bryce, 1965] *Bryce S. DeWitt* Dynamical theory of groups and fields. Gordon and Breach. N. Y., 1965. 248 p.
- [Garrett Lisi, 2007] *Garrett Lisi A.* An Exceptionally Simple Theory of Everything. URL: arXiv:0711.0770v1 [hep-th] 6 Nov. 2007.
- [Wesseling, 2001] *Wesseling P.* Principles of Computational Fluid Dynamics. Springer Series in Computational Mathematics, 2001. 644 p.

THREE-DIMENSIONAL TENSOR MATHEMATICS OF COMPUTATIONAL EXPERIMENT IN FLUIDMECHANICS

A. Bogdanov, A. Degtyarev, V. Khrumushin

Saint-Petersburg State University (SPbSU)

Architecture of a digital computing system determines the technical foundation of a unified mathematical language for exact arithmetic–logical description of phenomena and laws of continuum mechanics for applications in fluid mechanics and physical field theory. Enhancement of modern computers involves computing operations in optimally formalized set of locally connected digital geometric objects, existing and transforming independently, and being surrounded by similar virtual entities. Deep parallelization of computing processes serves to the revival of application of functional programming at a new technological level, providing traceability of computing processes with automatic application of multiscale hybrid circuits and adaptive mathematical models for the true reproduction of the fundamental laws of physics and continuum mechanics. Tensor formalization of numerical objects and computing operations serves to spatial interpolation of rheological state parameters and laws of the fluid mechanics as mathematical models in the local coordinates of the elementary numeric cells — large liquid particles.

Keywords: computational experiment, tensor mathematics, fluid mechanics.

Bogdanov Alexander Vladimirovich — doctor of physical and mathematical sciences, professor, bogdanov@csa.ru

Degtyarev Alexander Borisovich — doctor of technical sciences, professor, deg@csa.ru

Khrumushin Vasily Nikolaevich — PhD, doctoral student, khram@mail.ru

ТРАНСПОРТНАЯ МОДЕЛЬ МОСКОВСКОГО РЕГИОНА

А. Э. Воробьёв, А. Ю. Титов, В. А. Гаврилин, А. Ю. Меньшутин, И. А. Бахирев

Научно-исследовательский и проектный институт Генерального плана города Москвы (НИИПИ Генплана Москвы)

Прогнозирование транспортной ситуации является одной из ключевых градостроительных задач. В статье приводится описание одного из возможных способов построения модели прогнозирования транспортной системы. Представлена практическая реализация предложенных алгоритмов моделирования на примере модели транспортной системы Московского региона, разрабатываемой специалистами института Генерального плана города Москвы.

Исследование выполнено в Институте прикладной математики им. М. В. Келдыша РАН (ИПМ им. М. В. Келдыша РАН) в рамках Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы», Соглашение № 14.604.21.0052 от 30.06.2014 г. с Минобрнаукой. Уникальный идентификатор проекта RFMEFI60414X0052.

Ключевые слова: компьютерное моделирование, статическая макроскопическая транспортная модель, транспортные потоки.

ВВЕДЕНИЕ

Сбалансированность транспортной и социальной инфраструктур представляется необходимым условием долгосрочного устойчивого территориального и социально-экономического развития территории. Для формирования сбалансированного плана развития города необходим надёжный инструмент оценки последствий реализации того или иного градостроительного мероприятия и их совокупности. Для оценки градостроительных мероприятий в области транспорта таким инструментом могут быть статические транспортные макромодели. Их особенность заключается в возможности моделирования транспортной ситуации на протяжённых участках транспортной сети (например, на улично-дорожной сети города, региона, страны или континента). При этом транспортный поток описывается в макроскопических моделях максимально упрощённо: для описания такого потока на каждом участке транспортной сети используется лишь несколько его макроскопических характеристик, например, скорость и интенсивность. Таким образом, автотранспортный поток считается равномерным и имеет постоянные характеристики на протяжении каждого участка дороги. Это существенно сокращает вычислительную сложность модели и, соответственно, время её работы, что позволяет использовать макромодели таких сложных транспортных систем, как Московский регион, на стандартных рабочих станциях.

Воробьёв Алексей Эдуардович — инженер, avorobyev@genplanmos.ru

Титов Алексей Юрьевич — инженер, atitov@genplanmos.ru

Гаврилин Владимир Алексеевич — инженер, vgavrilin@genplanmos.ru

Меньшутин Антон Юрьевич — начальник отдела, кандидат физико-математических наук, amenshutin@genplanmos.ru

Бахирев Игорь Александрович — руководитель, кандидат технических наук, ibakhirev@genplanmos.ru

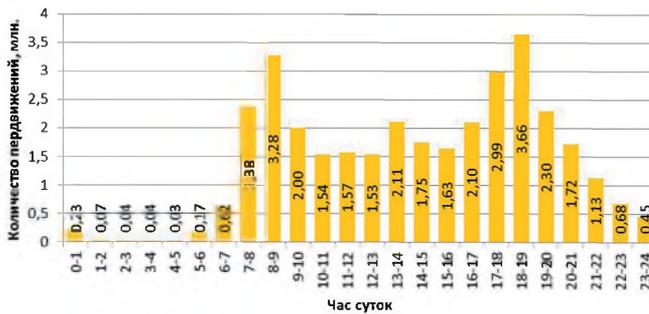


Рис. 1. Распределение передвижений жителей Московского региона по часам суток в рабочий день

ТРАНСПОРТНАЯ МОДЕЛЬ МОСКОВСКОГО РЕГИОНА

Для прогнозирования изменений транспортной ситуации в Москве и Московской области специалистами института Генерального плана города Москвы на базе программного комплекса Citilabs Cube разработана статическая макроскопическая транспортная модель Московского региона.

Для выявления существующих поведенческих характеристик населения был проведён ряд социологических опросов жителей Московского региона. С их помощью была определена целевая структура передвижений, характер пользования различными видами транспорта и другие параметры корреспонденций. Полученные данные были использованы при разработке и калибровке транспортной модели региона.

Разработанная модель описывает состояние транспортной системы на территории Москвы и Московской области в период пиковых нагрузок с 8 до 9 утра рабочего дня (рис. 1).

Статическая макроскопическая транспортная модель представляет собой математическую модель, построенную по принципу равновесия между спросом на транспортные услуги и транспортным предложением. Транспортная инфраструктура (автодороги, система общественного транспорта) формирует транспортное предложение, а желающие воспользоваться транспортными услугами люди определяют транспортный спрос.

ИСХОДНЫЕ ДАННЫЕ МОДЕЛИ

Транспортный спрос

Для описания распределения объектов, порождающих перемещения населения на территории Московского региона, область моделирования разбивается на N транспортных районов, каждому из которых присваивается порядковый номер от 1 до N (один транспортный район может соответствовать, например, микро-району). Каждый транспортный район i имеет набор параметров, определяющих спрос населения на транспортные услуги, связанные с выездом из района и с приездом в район: численность населения Pop_i , количество мест приложения труда Wrk_i и мощность «прочих» объектов Oth_i . Последний параметр характеризует район с точки зрения привлекательности для поездок в него с нетрудовыми целями (например, к торговым или рекреационным объектам).

Транспортная инфраструктура

Транспортная инфраструктура описывается ориентированным графом транспортной сети $\Gamma(V, E)$, где V — множество вершин, E — множество рёбер сети. Граф транспортной сети состоит из следующих частей:

1. Граф улично-дорожной сети (УДС).
В графе УДС каждое ребро соответствует участку дороги без перекрёстков. Направление ребра задаёт направление движения транспорта. Двухсторонние улицы описываются двумя рёбрами, направленными в разные стороны. Вершины графа улично-дорожной сети соответствуют перекрёсткам.
2. Граф внеуличного транспорта (метрополитена и железной дороги).
Рёбрам графа внеуличного транспорта соответствуют участки сети внеуличного транспорта между остановочными пунктами (например, перегон метрополитена между двумя станциями), а вершинам — остановочные пункты.
3. Вспомогательные рёбра, связывающие между собой граф внеуличного транспорта и граф улично-дорожной сети, а также рёбра, связывающие между собой различные части графа внеуличного транспорта (например, рёбра, соответствующие переходам между станциями метрополитена).
4. Множество вершин $Z = \{z_i, \text{ где } i = 1, \dots, N\}$, называемых *центроидами*, каждая из которых находится в геометрическом центре транспортного района с номером i . Центроиды — это пункты, порождающие и поглощающие транспортные потоки. Центроиды соединяются с графом сети специальными рёбрами.

Каждой паре районов отправления-прибытия $(z_i, z_j) \in Z \times Z$ соответствует спрос на перевозку w_{ij} — общий объём пользователей, которые должны прибыть из района отправления i в район прибытия j с 8 до 9 утра. Набор $(w_{ij} : (z_i, z_j) \in Z \times Z)$ называется матрицей корреспонденций. Транспортные услуги предоставляются населению тремя способами (по способам передвижения): на личном транспорте, на общественном транспорте и пешком. Пусть γ — показывающий способ передвижения, индекс, который может принимать значения *auto* (для передвижений на личном автотранспорте), *pt* (для передвижений на общественном транспорте) и *walk* (для пеших передвижений):

$$w_{ij} = w_{ij}^{auto} + w_{ij}^{pt} + w_{ij}^{walk} = \sum_{\gamma} w_{ij}^{\gamma} \quad (1)$$

Путём (маршрутом) в сети Γ , соединяющим вершины z_i и z_j , назовём последовательность рёбер

$$e_1 = (z_i \rightarrow v_1), e_2 = (v_1 \rightarrow v_2), \dots, e_m = (v_{m-1} \rightarrow v_m), e_{m+1} = (v_m \rightarrow z_j),$$

где $e_k \in E$ для всех $k = 1, \dots, m+1$ и $v_k \in V$ для всех $k = 1, \dots, m$. В маршрутах предполагается отсутствие петель и циклов. Обозначим через P_{ij} множество альтернативных маршрутов, соединяющих вершины z_i и z_j . Совокупность всех путей в сети Γ обозначим через P :

$$P = \bigcup_{z_i, z_j \in Z} P_{ij} \quad (2)$$

Перемещаясь между центроидами, пользователи сети выбирают тот или иной маршрут следования. Обозначим через x_p^{γ} поток или число пользователей,

переместившихся по пути $p \in P$ способом γ . Загрузка всей сети задаётся вектором $x^\gamma = (x_p^\gamma : p \in P)$.

Элементы матрицы корреспонденций и значения потоков по маршрутам связывает следующее соотношение:

$$w_{ij}^\gamma = \sum_{p \in P_{ij}} x_p^\gamma \quad (3)$$

Преодоление каждого из путей $p \in P$ сопровождается некоторыми затратами на перемещение (время, плата за проезд и т.д.) G_p^γ . В описываемой модели все затраты переводятся во время и измеряются в минутах. Важно отметить, что затраты на перемещение могут зависеть от загрузки транспортной сети в целом, а не только от потока по маршруту p , т.е. $G_p^\gamma = G_p^\gamma(x^\gamma)$. Считается, что затраты на перемещение по пути p складываются из затрат на прохождение рёбер графа, составляющих этот маршрут.

Обозначим через y_e^γ величину потока по ребру $e \in E$. Загрузка ребра складывается из потоков, движущихся по маршрутам, которые проходят по этому ребру:

$$y_e^\gamma = \sum_{p \in P} \theta_{ep} x_p^\gamma, \quad (4)$$

где

$$\theta_{ep} = \begin{cases} 1, & \text{если путь } p \text{ проходит через ребро } e, \\ 0, & \text{в противном случае.} \end{cases}$$

Затраты на прохождение единицы потока по дуге e обозначим через τ_e^γ . В модели предполагается, что значение τ_e^γ однозначно определяется величиной потока y_e^γ . При этом существует ненулевые затраты на прохождение ребра по свободной сети $\tau_e^\gamma(0) > 0$ (если по ребру запрещено перемещение способом γ , то можно считать $\tau_e^\gamma(0) = \infty$).

Тогда затраты на перемещение по пути p можно найти по формуле:

$$G_p^\gamma = \sum_{e \in E} \theta_{ep} \tau_e^\gamma(y_e^\gamma). \quad (5)$$

Обозначим через c_{ij}^γ затраты на перемещение способом γ из района i в район j :

$$c_{ij}^\gamma = \frac{\sum_{p \in P_{ij}} G_p^\gamma x_p^\gamma}{w_{ij}^\gamma}. \quad (6)$$

Будем считать, что усреднённые затраты на перемещение из района i в район j являются средними по всем способам перемещения:

$$c_{ij} = \frac{\sum_{\gamma} w_{ij}^\gamma c_{ij}^\gamma}{w_{ij}}. \quad (7)$$

Алгоритмы работы модели

В основе модели Московского региона лежит пошаговый алгоритм поиска равновесного состояния транспортной системы, который состоит из следующих четырёх шагов:

1. **Trips Generation.** Вычисление для каждого транспортного района объёмов прибытия и отправления пользователей сети. Объёмы прибытия–отправления определяются на основании статистических данных о закономерностях поведения населения.
2. **Trips Distribution.** Вычисление матрицы корреспонденций. Матрица определяется на основании данных о затратах на перемещение между транспортными районами, а также об объёмах прибытия и отправления.
3. **Mode Choice.** Для каждой пары районов прибытия–отправления вычисляются вероятности использования для перемещения между районами одного из трех способов передвижения. Вероятность выбора способа передвижения зависит от соотношения затрат на перемещения между каждым из трёх способов. Таким образом, матрица корреспонденций разделяется (расщепляется) на три слагаемых.
4. **Traffic Assignment.** Распределение корреспонденций для каждого способа перемещения по маршрутам, определение затрат на перемещения для каждой пары районов прибытия–отправления.

Распределение объёмов прибытия–отправления в матрицу корреспонденций, которое происходит на втором шаге модели, зависит от затрат на передвижение, которые в свою очередь будут зависеть от числа пользователей, желающих воспользоваться услугами транспортной сети. Модель позволяет найти состояние равновесия между этими двумя взаимозависимыми объектами. Ниже более подробно описаны алгоритмы, использующиеся в модели на каждом шаге.

Trips Generation

При проведении социологических исследований установлены закономерности поведения жителей Московского региона, которые позволяют на основании параметров, характеризующих транспортный район i (Pop_i , Wrk_i и Oth_i), определить общее число прибывающих в район D_i и убывающих из района O_i человек за период с 8 до 9 часов. Всевозможные перемещения жителей Московского региона разделяются на так называемые слои спроса. Каждый слой спроса характеризуется объектом отправления и объектом прибытия. Например, трудовые корреспонденции населения из дома на работу относятся к слою спроса дом–работа. В описываемой модели выделено 3 типа объектов прибытия–отправления (дом, работа и прочее), для которых может быть сформировано 9 слоёв спроса. Следует отметить, что для некоторых слоёв спроса (например, дом–дом) число поездок в утренний период мало, поэтому такие поездки обычно исключаются из рассмотрения. Число поездок в слое спроса определяется коэффициентом подвижности — числом поездок от объектов отправления к объектам прибытия на тысячу жителей, которые совершают пользователи сети в течение периода моделирования (с 8 до 9 часов). Примеры коэффициентов подвижности, которые используются в модели для определения целей перемещения населения в утренний час пик, приведены в табл. 1. В строках указан тип объекта отправления, а в столбцах — тип объекта прибытия.

Наибольшее число поездок в рассматриваемый час происходит из дома на работу. Объёмы прибытия D_i и отправления O_i в утренний час пик определяются по формулам:

$$\begin{cases} O_i = 0,130Pop_i + 0,029Pop_i + 0,004Wrk_i + 0,005Oth_i + 0,004Oth_i, \\ D_i = 0,004Pop_i + 0,005Pop_i + 0,013Wrk_i + 0,004Wrk_i + 0,029Oth_i. \end{cases} \quad (8)$$

Таблица 1. Значение коэффициентов подвижности населения Московского региона с 8 до 9 часов утра

	Дом	Работа	Прочее
Дом	0,000	0,130	0,029
Работа	0,004	0,000	0,000
Прочее	0,005	0,004	0,000

Trips Distribution

Предполагается, что объёмы прибытия отправления распределяются между районами прибытия-отправления согласно правилу (9), которое в литературе получило название гравитационной модели [Введение в мат. моделирование..., 2013]:

$$w_{ij} = \frac{O_i D_j}{e^{-\beta c_{ij}}}, \quad i, j = 1, \dots, N, \quad (9)$$

$$\left. \begin{aligned} O_i &= \sum_{j=1}^N w_{ij}, \\ D_j &= \sum_{i=1}^N w_{ij}, \\ i, j &= 1, \dots, N. \end{aligned} \right\} \quad (10)$$

Знаменатель в формуле (9) — гравитационная функция; параметр β — константа (коэффициент гравитационной функции), которая определяется в процессе калибровки модели. Для Московского региона значение β составляет 0,08. Уравнения (10) называются балансовыми ограничениями.

На практике система (9)–(10) обычно решается с использованием специального итерационного алгоритма [Введение в мат. моделирование..., 2013]. Для процесса поиска равновесия указанная система является ключевой, так как именно она определяет взаимосвязь между затратами на перемещение и объёмами корреспонденций.

Mode Choice

Определение способа передвижения между парой районов прибытия-отправления происходит согласно модели дискретного выбора [Ortuzar, Willumsen, 2011]. В работах Т. Доменкича и Д. Мак-Фаддена [Domencich, McFadden, 1975] и Х. Вильямса [Williams, 1977] показано, что если пользователи действуют рационально (т. е. выбирают способ передвижения с наименьшими затратами) в условиях наличия неполной информации о системе, то вероятность выбора способа передвижения γ между районами отправления i и прибытия j будет определяться формулой:

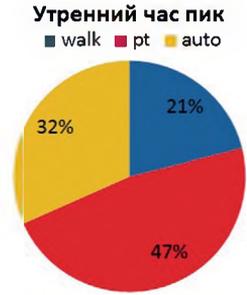
$$\rho_{ij}^{\gamma} = \frac{e^{-\lambda c_{ij}^{\gamma}}}{\sum_{\gamma} e^{-\lambda c_{ij}^{\gamma}}}. \quad (11)$$

Здесь λ — константа (коэффициент модели дискретного выбора) и определяется в процессе калибровки модели. Для Московского региона значение λ составляет 0,08.

Рис. 2. Распределение корреспонденций по способам передвижения в утренний час пик. Результаты моделирования

Таким образом, на третьем шаге транспортной модели матрица корреспонденций расщепляется на три слагаемых по способам передвижения (рис. 2):

$$w_{ij}^Y = \rho_{ij}^Y w_{ij}. \quad (12)$$



Traffic Assignment

Распределение корреспонденций по маршрутам происходит независимо для каждого способа передвижения. Рассмотрим каждый из них отдельно.

Пешие передвижения

Для пеших корреспонденций затраты на передвижения не зависят от уровня загрузки сети, т. е.

$$\tau_e^{walk}(y_e^{walk}) = \tau_e^{walk}(0). \quad (13)$$

Все пользователи выбирают маршруты передвижения обладающие минимальными затратами на перемещение из всех возможных вариантов. Таким образом, задача распределения пеших корреспонденций по маршрутам сводится к задаче поиска кратчайшего пути во взвешенном графе.

Передвижения на личном автотранспорте

Для автомобильных корреспонденций затраты на передвижение по ребру графа e зависят от потока по этому ребру y_e^{auto} следующим образом:

$$\tau_e^{auto}(y_e^{auto}) = \tau_e^{auto}(0) \left[1 + \alpha \left(\frac{y_e^{auto}}{C_e} \right)^\mu \right], \quad (14)$$

где $\tau_e^{auto}(0)$, α , μ и C_e являются параметрами характеризующими участок дороги (ребро графа). Такая зависимость в транспортном моделировании известна, как BPR-функция (Bureau of Public Road). Параметр C_e соответствует так называемой «пропускной способности» участка УДС, которая в зависимости от типа дороги варьируется от 600 до 2000 автомобилей в час на одну полосу движения. Величину y_e^{auto}/C_e будем называть коэффициентом загрузки ребра (участка УДС).

Распределение автомобильных корреспонденций по сети происходит в соответствии с правилом, получившим название первого принципа Вардропа [Wardrop, 1952]: «Пользователи сети независимо друг от друга выбирают маршруты следования, соответствующие их минимальным затратам на передвижение». Во введённых обозначениях первый принцип Вардропа можно формализовать следующим образом: если по пути $p \in P_{ij}$ идёт ненулевой поток \hat{x}_p^{auto} , то затраты по этому пути минимальны:

$$\text{если } \hat{x}_p^{auto} > 0, \text{ то } G_p^{auto}(\hat{x}^{auto}) = \min_{q \in P_y} G_q^{auto}(\hat{x}^{auto}), \quad (15)$$

где загрузка сети определяется вектором \hat{x}^{auto} . Допустимой областью значений этого вектора является множество:

$$X^{auto} = \left\{ x_p^{auto} \geq 0 \text{ для всех } p \in P : \sum_{p \in P_y} x_p^{auto} = w_{ij}^{auto} \right\}.$$

Задача поиска вектора \hat{x}^{auto} эквивалентна решению вариационного неравенства [Введение в мат. моделирование..., 2013]:

$$G(\hat{x}^{auto})(x^{auto} - \hat{x}^{auto}) \geq 0 \forall x^{auto} \in X^{auto}, \text{ где } G(x) = \left\{ G_p^{auto}(x) : p \in P \right\}. \quad (16)$$

При условии (14) неравенство (16) сводится к оптимизационной задаче:

$$\sum_{e \in E} \int_0^{y_e^{auto}} \tau_e^{auto}(z) dz \rightarrow \min, \quad y_e^{auto} = \sum_{p \in P} \theta_{ep} x_p^{auto}, \quad x^{auto} \in X^{auto}. \quad (17)$$

Существующие методы численного решения полученной оптимизационной задачи описаны в пособии под редакцией А. В. Гасникова [Введение в мат. моделирование..., 2013].

Передвижения на пассажирском транспорте общего пользования

Для поездок на общественном транспорте затраты на передвижение по ребру графа e , в том случае, если по ребру e разрешено движение на автомобиле, определяются затратами на перемещение по этому ребру на автомобиле:

$$\tau_e^{pt}(y_e^{pt}) = \min \left\{ \tau_e^{pt}(0), \tau_e^{auto}(y_e^{auto}) \right\}. \quad (18)$$

В случае, если по ребру e запрещено движение автомобилей, то затраты на передвижение по этому ребру на общественном транспорте не зависят от потока по ребру:

$$\tau_e^{pt}(y_e^{pt}) = \tau_e^{pt}(0). \quad (19)$$

В дополнение к графу транспортной сети транспортную инфраструктуру для перемещений на общественном транспорте составляет информация о линиях пассажирского транспорта: остановочные пункты, интервалы движения, подвижной состав и т.д. Поэтому поиск кратчайшего пути (т.е. пути обладающего минимальными затратами на передвижение) происходит в специальном образом модифицированном графе транспортной сети. Более подробно с алгоритмами распределения потоков читатель может ознакомиться в работе Дж. Ортузара и Л. Виллимсена [Ortuzar, Willumsen, 2011].

Процесс поиска равновесного состояния

Процесс поиска равновесного состояния транспортной системы Московского региона происходит итерационно. В итерации можно выделить два этапа (рис. 3):

- вспомогательный цикл Mode Equilibrium, в котором для фиксированной матрицы корреспонденций (w_{ij}) происходит поиск равновесного распределение корреспонденций по способам передвижения $\left\{ w_{ij}^y \right\}$;

Рис. 3. Схема поиска равновесного состояния транспортной системы

- основной цикл Distribution Equilibrium, в котором устанавливается равновесие между общим спросом (w_{ij}) и затратами на передвижение (c_{ij}).

На первой итерации во втором шаге (Trips Distribution) для вычисления значений w_{ij} используются затраты на передвижение c_{ij} по свободной транспортной сети, т.е. затраты на ребре e равны $\tau_e^i(0)$. На второй и последующих итерациях основного цикла во втором шаге для вычисления значений w_{ij} используются затраты на передвижение c_{ij} с предыдущей итерации. Процесс поиска равновесного состояния транспортной системы происходит до тех пор, пока не будет выполнено неравенство:

$$\sum_{i=1}^N \sum_{j=1}^N |w_{ij}^{(n)} - w_{ij}^{(n-1)}| \leq \epsilon_w, \quad (20)$$

где $w_{ij}^{(n)}$, $w_{ij}^{(n-1)}$ — значения элементов матрицы корреспонденций на n и $(n - 1)$ итераций соответственно, а значение параметра ϵ_w определяется условиями задачи. Экспериментально было установлено, что для модели Московского региона в большинстве случаев достаточно провести 5 итераций основного цикла для установления равновесия.

Затраты на передвижение по свободной сети c_{ij}^γ также используются в первой итерации во вспомогательном цикле на третьем шаге (Mode Choice) для вычисления w_{ij}^γ . На последующих итерациях цикла Mode Equilibrium используются затраты на передвижение с предыдущего цикла. Процесс поиска равновесного расщепления корреспонденций по способам передвижения происходит до тех пор, пока не будет выполнено неравенство:

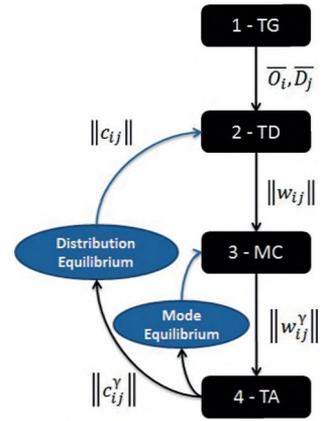
$$\sum_{\gamma} \sum_{i=1}^N \sum_{j=1}^N |c_{ij}^{\gamma(n)} - c_{ij}^{\gamma(n-1)}| \leq \epsilon_c, \quad (21)$$

где $c_{ij}^{\gamma(n)}$, $c_{ij}^{\gamma(n-1)}$ — затраты на перемещение способом γ на n и $(n - 1)$ итерации соответственно, а значение параметра ϵ_c определяется условиями задачи. Экспериментально было установлено, что для модели Московского региона в большинстве случаев достаточно провести 3 итерации вспомогательного цикла для установления равновесия.

Таким образом, для проведения одного расчёта необходимо выполнить наиболее дорогую с вычислительной точки зрения часть модели — шаг Traffic Assignment — не менее 15 раз. Время работы модели на рабочей станции составляет в зависимости от условий задачи 5–10 часов.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Результатом работы модели является набор показателей, характеризующих состояние транспортной системы, в описываемый период времени:



- число передвижений w_{ij} и w_{ij}^y ;
- данные о затратах на передвижение — c_{ij} и c_{ij}^y ;
- распределение корреспонденций по маршрутам — x_p^y для всех $p \in P_{ij}$.



Рис. 4. Транспортные потоки на рёбрах графа УДС. Толщина линии пропорциональна загрузке ребра. Результаты моделирования

На основании этих данных по формуле (4) могут быть вычислены транспортные потоки на рёбрах графа сети u_e^y (рис. 4 и 5).

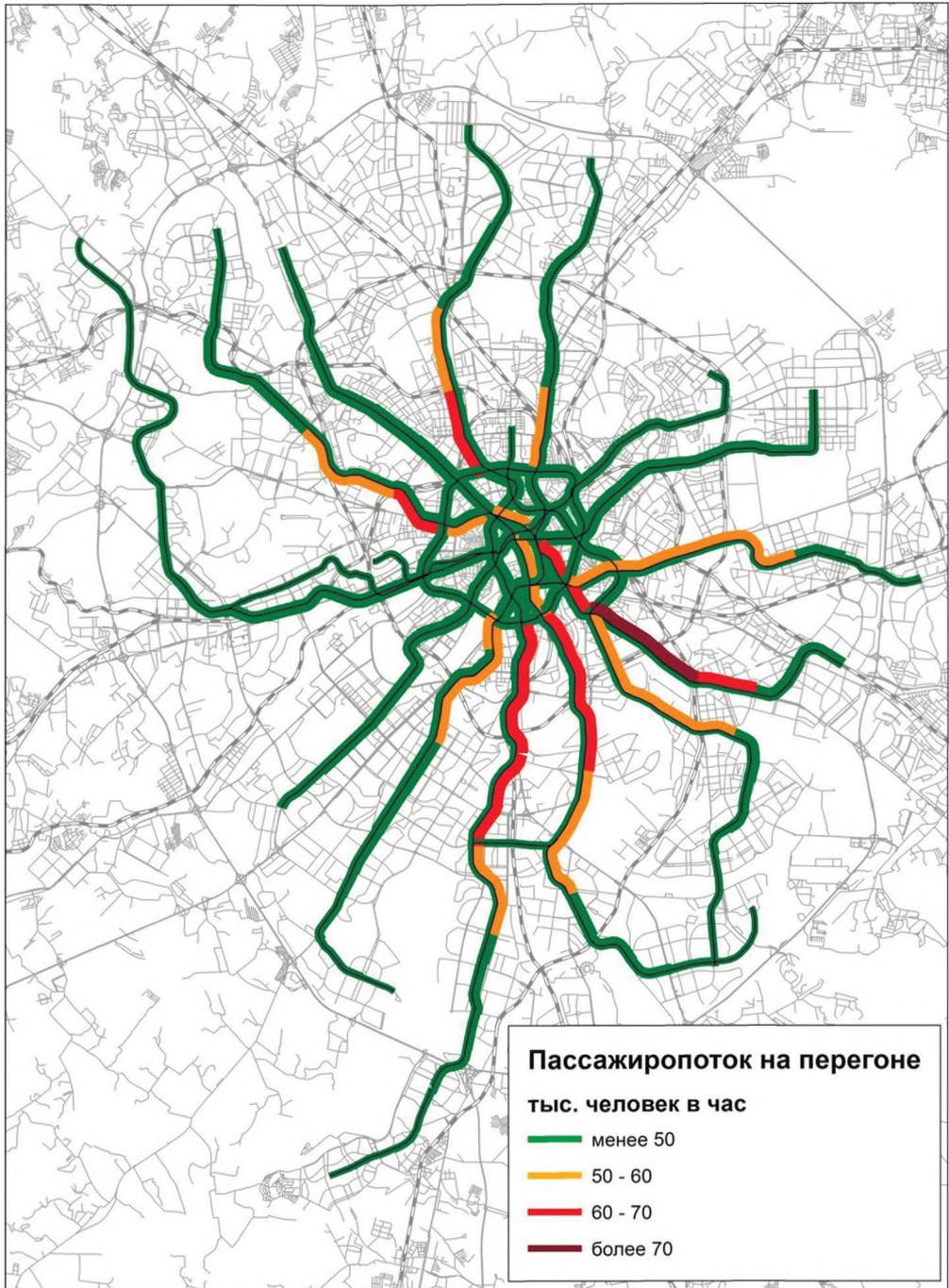


Рис. 5. Пассажиропотоки Московского метрополитена. Результаты моделирования

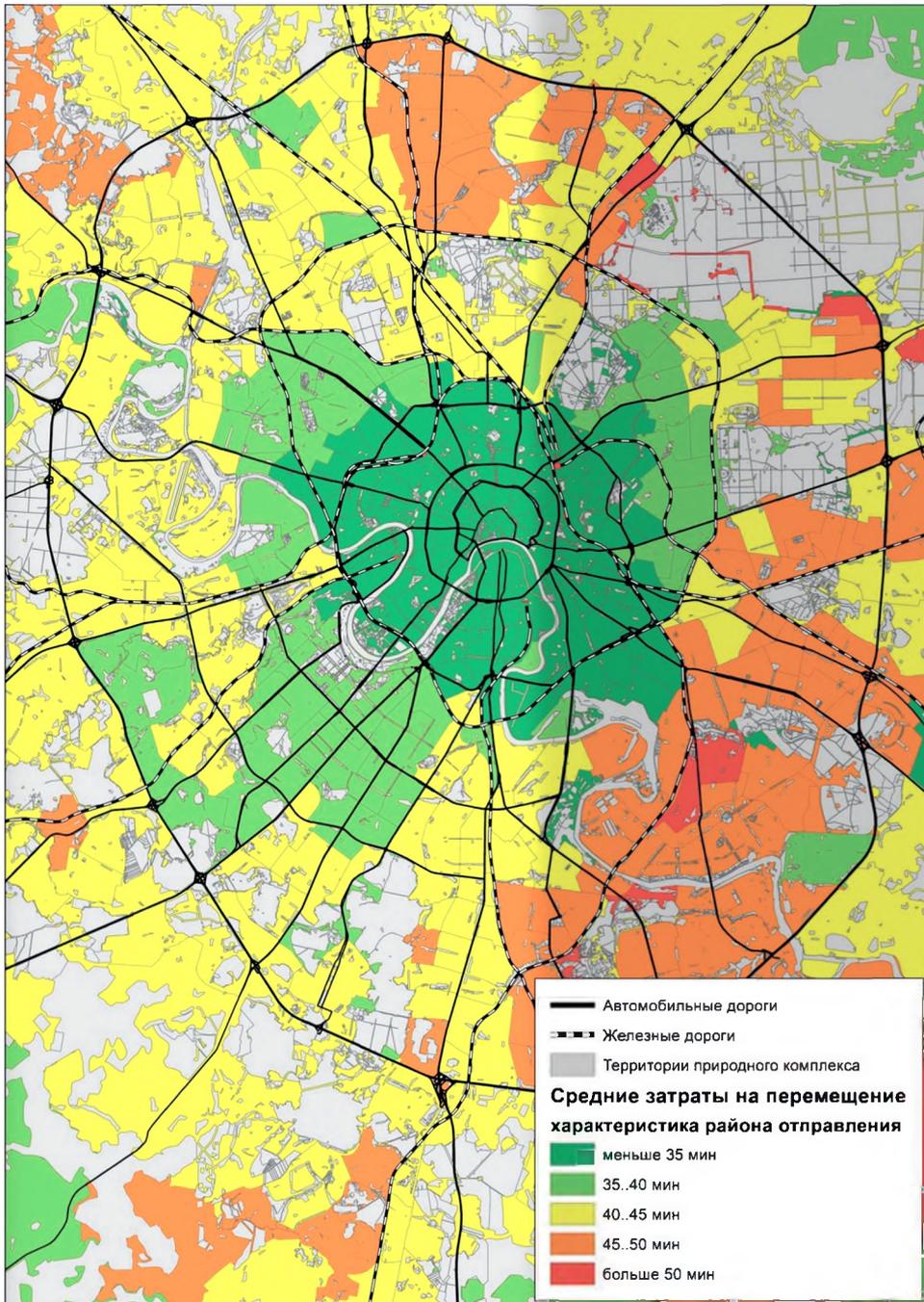


Рис. 6. Средние затраты на перемещения агрегированные по районам отправления. Результаты моделирования

При моделировании текущего состояния транспортной системы, потоки на рёбрах графа сравниваются с реальными данными о транспортных или пассажирских потоках на соответствующих участках УДС или перегонах ско-

ростного внеуличного транспорта. К источникам реальных данных о существующей загрузке транспортной инфраструктуры следует отнести натурные обследования и различные автоматизированные системы сбора информации, такие как турникеты на входах на станции Московского метрополитена, системы АСКП (Автоматизированная система контроля проезда) и АСОКУПЭ (Автоматизированная система оплаты контроля и учёта проезда в пригородных электропоездах), датчики движения транспорта, GPS- и ГЛОНАСС-треки общественного транспорта.

Число передвижений и затраты могут быть агрегированы по району прибытия или отправления (рис. 6, полученные показатели будем называть характеристиками транспортного района прибытия или отправления), а также по всей сети (полученные показатели будем называть общесетевыми показателями). В табл. 2 приведено сравнение общесетевых показателей полученных в результате моделирования текущего состояния транспортной системы с данными социологических опросов жителей Москвы и Московской области.

Таблица 2. Сравнение общесетевых показателей по результатам работы модели с данными социологических опросов населения

Показатель		Соцопрос	Модель
Количество передвижений [млн чел]	Пешие передвижения	0,68±0,07	0,55
	Общественный транспорт	1,52±0,09	1,69
	Личные автомобили	1,02±0,07	1,04
Объёмы миграции населения [млн чел]	Из Москвы в область	0,12±0,02	0,13
	Из области в Москву	0,28±0,03	0,29
Средние затраты на передвижение [мин]	Пешие передвижения	17,7±1,0	19,4
	Общественный транспорт	57,5±1,4	52,4
	Личные автомобили	40,0±2,0	38,7
Среднее затраты на передвижение по регионам прибытия-отправления [мин]	Из Москвы в Москву	42±5	44
	Из Москвы в область	75±15	54
	Из области в Москву	90±2	73
	Из области в область	30±2	28
Среднее расстояние поездки (по прямой) [км]	Общественный транспорт	12,7±1,2	12,8
	Личные автомобили	8,8±1,4	10,0
Количество пассажиров на различных видах общественного транспорта [млн чел]	Московский метрополитен	0,96±0,07	1,06
	Пригородные электропоезда	0,23±0,02	0,48
	НГПТ	1,12±0,15	0,85

ПРОГНОЗИРОВАНИЕ ИЗМЕНЕНИЙ ТРАНСПОРТНОЙ СИТУАЦИИ

Моделирование позволяет определить не только существующие значения показателей, характеризующих состояние транспортной системы, но и их изменение в случае реализации различных влияющих на транспортную ситуацию мероприятий. Для этого необходимо внести соответствующие изменения в исходные данные модели. Мероприятия, затрагивающие транспортную систему, могут влиять как на спрос (создание новых мест приложения труда или жилья — изменение

параметров транспортных районов), так и на предложение (строительство новых или реконструкция существующих автодорог, внесение изменений в систему общественного транспорта — изменение графа транспортной сети). Разработанная модель используется как для долгосрочных, так и для краткосрочных прогнозов, и позволяет оценивать изменение транспортной обстановки не только во всей транспортной сети, но и на отдельных её участках.

ЛИТЕРАТУРА

- [Введение в мат. моделирование..., 2013] Введение в математическое моделирование транспортных потоков: учеб. пособие / Под ред. Гасникова А.В.; Авт. кол. *Гасников А. В., Кленов С.Л.* и др.; 2-е изд. испр. и дополн. М.: МЦНМО, 2013. 428 с.
- [Domencich, McFadden, 1975] *Domencich T., McFadden D.* Urban Travel Demand: A Behavioural Analysis. Amsterdam: North-Holland, 1975.
- [Ortuzar, Willumsen, 2011] *Ortuzar J. de Dios, Willumsen L. G.* Modelling Transport. UK: John Wiley and Sons, Ltd, 2011. 586 p.
- [Wardrop, 1952] *Wardrop J.* Some Theoretical Aspects of Road Traffic Research // Proc. Inst. Civil Engineers. 1952. V. 1. P. 325–362.
- [Williams, 1977] *Williams H. C. W. L.* On the formation of travel demand models and economic evaluation measures of user benefit // Environment and Planning 1977. V. 9A. P. 285–344.

TRANSPORT MODEL OF MOSCOW REGION

A. E. Vorobyev, A. Y. Titov, V. A. Gavrilin, A. Y. Menshutin, I. A. Bakhirev

R&P Institute of Moscow City Master Plan

Prediction of the transport situation is one of the key urban problems. In this paper, one of the possible ways to build a predictive model of the transport system is described. A practical implementation of the proposed algorithms for simulation Moscow region transport system is presented.

Keywords: computer simulation, static macroscopic transport model, traffic flows.

Vorobyev Aleksey Eduardovich — engineer, avorobyev@genplanmos.ru

Titov Aleksey Yuryevich — engineer, atitov@genplanmos.ru

Gavrilin Vladimir Alekseevich — engineer, vgavrilin@genplanmos.ru

Menshutin Anton Yuryevich — head of department, PhD, amenshutin@genplanmos.ru

Bakhirev Igor Aleksandrovich — head of department, PhD, ibakhirev@genplanmos.ru

НЕКЛАССИЧЕСКИЕ РЕЛАКСАЦИОННЫЕ КОЛЕБАНИЯ И БИФУРКАЦИЯ ГОЛУБОГО НЕБА В ЗАДАЧЕ МОДЕЛИРОВАНИЯ КАРДИОРИТМОВ

С. Д. Глызин¹, А. Ю. Колесов¹, Н. Х. Розов²

¹ Ярославский государственный университет им. П. Г. Демидова (ЯрГУ)

² Московский государственный университет имени М. В. Ломоносова (МГУ)

Рассматривается новый класс систем обыкновенных дифференциальных уравнений с одной быстрой и двумя или более медленными переменными. В работе показано, что в случае выполнения некоторых условий технического характера на функции правых частей модели, в ней могут одновременно реализовываться так называемые неклассические релаксационные колебания и бифуркация голубого неба. Под неклассическими релаксационными колебаниями понимаются колебания, у которых медленные компоненты асимптотически близки к некоторым разрывным по времени функциям, а быстрая компонента δ -образна. В работе так же построено отображение, отвечающее за аттрактор системы (отображение Пуанкаре вдоль траекторий на специальном образе выбранном сечении), с помощью которого удаётся обосновать наличие указанной бифуркации. Разбираются случаи, когда в результате катастрофы голубого неба возникает устойчивый релаксационный цикл или устойчивый двумерный инвариантный тор. Рассматривается также вопрос о появлении в системе гиперболического аттрактора (аттрактора Смейла-Вильямса). Сделана попытка применения обобщаемых систем к моделированию кардиоритмов.

Исследование выполнено при поддержке Российского научного фонда (проект № 14-21-00158).

Ключевые слова: сингулярно возмущённая система, релаксационный цикл, асимптотика, устойчивость, катастрофа голубого неба, моделирование кардиоритмов.

ВВЕДЕНИЕ

Описание функционирования сердечной мышцы в первую очередь связано с задачей математического моделирования его электрической активностью. Как известно, кардиосигнал представляет собой сигнал нестационарной и негармонической природы. Данный сигнал зависит от индивидуальных биологических свойств конкретного человека и состоит из зубцов, интервалов и сегментов. Зубцы кардиосигнала — это локальные экстремумы графика ЭКГ (электрокардиограмма). Сегментом кардиосигнала называют участок упомянутого графика между соседними зубцами, а интервал кардиосигнала состоит из зубца (комплекса зубцов) и сегмента. Типовой график ЭКГ представлен на рис. 1 (данный график заимствован из статьи П. Е. МакШэрри с соавторами [McSharry et al., 2003]). Согласно этому графику можно выделить следующие основные фрагменты кардиограммы (см. [Казаков, 2012; McSharry et al., 2003]).

1) *P*-зубец, отражающий сокращение предсердий сердца (предсердную деполяризацию). Деполяризация начинается в клетках-водителях ритма синусового узла сердца, а затем распространяется по проводящим пучкам к правому и левому предсердию;

Глызин Сергей Дмитриевич — заведующий кафедрой, доктор физико-математических наук, профессор, glyzin@uniyar.ac.ru

Колесов Андрей Юрьевич — доктор физико-математических наук, профессор, kolosov@uniyar.ac.ru

Розов Николай Христович — декан, доктор физико-математических наук, профессор, член-корреспондент РАЕН, fro.mgu@mail.ru

2) *PQ*-интервал, отражающий прохождение импульса возбуждения по предсердиям сердца и атриовентрикулярному соединению до миокарда желудочков сердца;

3) *QRS*-комплекс, соответствующий сокращению желудочков сердца. Данный комплекс, который состоит из трёх зубцов *Q*, *R*, *S* и представляет сумму потенциалов деполяризующихся кардиомиоцитов внутренних и наружных слоёв миокарда;

4) *QT*-интервал, отражающий сумму процессов деполяризации и последующей реполяризации миокарда желудочков сердца. Этот интервал принято называть электрической систолой сердца;

5) *ST*-сегмент, соответствующий интервалу времени, когда миокард желудочков сердца полностью охвачен возбуждением. Он предшествует последней фазе сердечного цикла, в течение которой происходит восстановление сердечной мышцы после её сокращения;

6) *T*-зубец, представляющий период времени расслабления миокарда желудочков сердца. В этом периоде сердечная мышца находится в покое;

7) *RR*-интервал, т.е. полный сердечный цикл (время между соседними *P*-зубцами).

Принцип моделирования кардиоритмов, используемый многими авторами, состоит в подборе некоторой конкретной системы обыкновенных дифференциальных уравнений, в которой график зависимости от времени одной из компонент близок по форме к показанному на рис. 1. В частности, на этом пути было предложено несколько трёхмерных моделей. Одна из самых популярных рассматривается в статье [McSharry et al., 2003]. Далее, следует отметить так называемую модель Зимана [Zeeman, 1972], а также трёхмерную модель из статьи Н. Б. Янсона, А. Н. Павлова и В. С. Анищенко [Janson et al., 1998].

Существуют и многомерные модели сердечной активности, представляющие собой системы связанных уравнений Фитц-Хью-Нагумо, Ван дер Поля и т. д. (см., например, [dos Santos et al., 2004; Podziemski, Zebrowski, 2013; Suchorsky, Rand, 2009]).

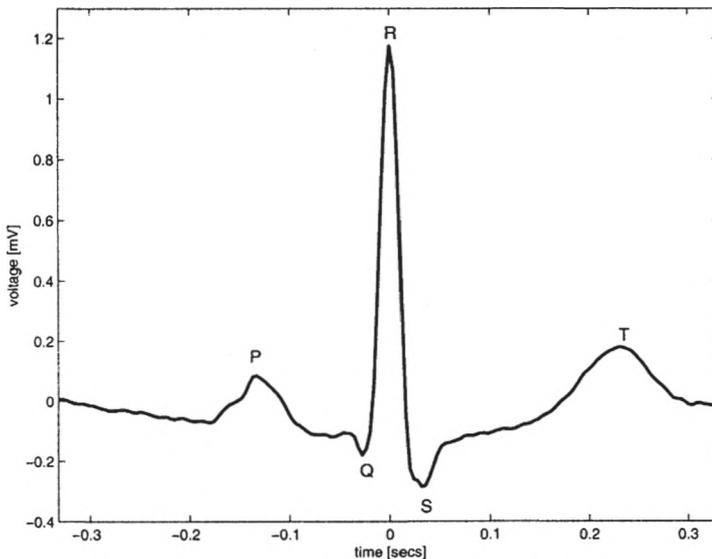


Рис. 1. Типовой график ЭКГ (см. [McSharry et al., 2003])

Наш подход к моделированию ЭКГ основан на синтезе двух идей: неклассических релаксационных колебаний и катастрофы голубого неба. В результате получается целый класс систем вида:

$$\begin{cases} \dot{x} = f_1(x, \mu) + \varepsilon^{p_1} y^{p_2} f_2(x), \\ \varepsilon \dot{y} = g(x) - h(y), \end{cases} \quad (1)$$

которые можно использовать как феноменологические модели электрической активности сердца. Способу определения функций f_1, f_2, g и h будет посвящён следующий раздел статьи, а показатели $0 < p_1 < p_2$ могут быть выбраны любыми. Система (1) была рассмотрена в статье [Глызин и др., 2015а] при $p_1 = 1/2, p_2 = 2$, где показана реализуемость в ней бифуркации голубого неба. В настоящей работе показывается, что в системе (1) при условиях, аналогичных приведённым в работе [Глызин и др., 2015б], также происходит указанная бифуркация. За счёт подбора параметров p_1, p_2 можно изменить порядок максимальных значений быстрой переменной y , отвечающей в нашем случае за электрическую активность сердечной мышцы, и получить решения системы (1), хорошо согласующиеся с наблюдаемыми графиками ЭКГ.

1. КАТАСТРОФА ГОЛУБОГО НЕБА В ТРЁХМЕРНОЙ СИСТЕМЕ С ДВУМЯ МЕДЛЕННЫМИ И ОДНОЙ БЫСТРОЙ ПЕРЕМЕННЫМИ В СЛУЧАЕ НЕКЛАССИЧЕСКИХ РЕЛАКСАЦИОННЫХ КОЛЕБАНИЙ

1.1. Постановка задачи

Прежде всего, скажем несколько слов о сути катастрофы голубого неба. Этим термином принято называть нелокальную бифуркацию коразмерности один, которая в простейшем случае состоит в следующем.

Рассмотрим гладкое однопараметрическое семейство векторных полей X_μ в \mathbb{R}^3 и предположим, что при $\mu = 0$ поток X_μ имеет периодическую траекторию L_0 типа простой седло-узел. Далее рассмотрим, некоторую достаточно малую окрестность \mathcal{U} траектории L_0 , разделяемую двумерным сильно устойчивым многообразием $W^{ss}(L_0)$ на две области: узловую \mathcal{U}^+ , все траектории из которой стремятся к L_0 при $t \rightarrow +\infty$, и седловую \mathcal{U}^- , в которой лежит двумерное неустойчивое многообразие $W_{loc}^u(L_0)$ с краем L_0 . Следующее ограничение носит существенно нелокальный характер и состоит в том, что $W_{loc}^u(L_0)$ при увеличении t сначала покидают окрестность \mathcal{U} , а затем снова возвращаются в неё, попадая в узловую область \mathcal{U}^+ . Тогда, очевидно, каждая из упомянутых траекторий оказывается двоякоасимптотической к L_0 . И наконец, будем считать, что множество $W^u(L_0)$, получающееся из $W_{loc}^u(L_0)$ после продолжения по траекториям потока X_0 , не является топологическим многообразием (в трёхмерном случае это означает, что его замыкание не гомеоморфно двумерному тору).

Как показано в статье [Гураев, Шильников, 1995], при сформулированных ограничениях и при некоторых дополнительных условиях технического характера исчезновение в системе X_μ , $0 < \mu \ll 1$ седло-узловой цикла L_0 приводит к появлению устойчивой замкнутой траектории $L(\mu)$, период и длина которой стремятся к бесконечности при $\mu \rightarrow 0$. Сама же траектория $L(\mu)$ имеет своим верхним топологическим пределом при $\mu \rightarrow 0$ множество $W^u(L_0) \cup L_0$. Описанная бифуркация получила название «катастрофа голубого неба».

В работах [Shilnikov et al., 2003; 2005] проиллюстрирована реализуемость упомянутой выше бифуркации в сингулярно возмущённых системах с одной

медленной и m , $m \geq 2$ быстрыми переменными. Далее, в статье [Глызин и др., 2008] катастрофа голубого неба изучалась в системе

$$\begin{cases} \dot{x} = f(x, y, \mu), \\ \varepsilon \dot{y} = g(x, y), \end{cases} \quad (2)$$

где $x = (x_1, x_2) \in \mathbb{R}^2$, $y \in \mathbb{R}$, $0 < \varepsilon \ll 1$, $|\mu| \ll 1$, а на функции $f, g \in C^\infty$ были наложены стандартные ограничения (см. [Мищенко, Розов, 1975]), обеспечивающие существование так называемых классических релаксационных колебаний. Напомним, что классическими называются колебания, у которых при $\varepsilon \rightarrow 0$ медленные компоненты x_1, x_2 стремятся к некоторым непрерывным по t функциям, а быстрая компонента y сходится поточечно к разрывной функции.

В настоящей статье результаты из работы [Глызин и др., 2008; 2015a] распространяются на аналогичную (2) трёхмерную систему (1), в которой, как и в (2), полагаем $x = (x_1, x_2) \in \mathbb{R}^2$, $y \in \mathbb{R}$, $0 < \varepsilon \ll 1$, $|\mu| < \mu_0$, а $\mu_0 > 0$ — некоторая достаточно малая константа. Что же касается функций $f_1(x, \mu) \in C^\infty(\mathbb{R}^2 \times [-\mu_0, \mu_0]; \mathbb{R}^2)$, $f_2(x) \in C^\infty(\mathbb{R}^2, \mathbb{R}^2)$, $g(x) \in C^\infty(\mathbb{R}^2)$, $h(y) \in C^\infty(\mathbb{R})$, то они будут удовлетворять специальным условиям, гарантирующим реализуемость неклассических релаксационных колебаний. Согласно принятой в книге [Мищенко и др., 1995] терминологии таковыми называются колебания, у которых при $\varepsilon \rightarrow 0$ медленные компоненты сходятся поточечно к разрывным функциям, а быстрая компонента является δ -образной.

Приступим к детальному описанию ограничений на правые части системы (1). Первые два из них связаны с топологией поверхности медленных движений

$$\Gamma = \{(x, y) : g(x) - h(y) = 0\}. \quad (3)$$

Условие 1. Предполагаем, что, во-первых, $g(x) > 0$, $\forall x \neq 0$, $g(0) = 0$; во-вторых, при любом фиксированном $z > 0$ уравнение $g(x) = z$ определяет в \mathbb{R}^2 замкнутую кривую класса C^∞ , гомеоморфную окружности.

Условие 2. Найдётся такое $y = y_* > 0$, что $h'(y) > 0$ при $y < y_*$, $h'(y) < 0$ при $y > y_*$, $h'(y_*) = 0$, $h''(y_*) < 0$, $h(y_*) = z_* > 0$. Считаем также, что $h(0) = 0$, а при $y \rightarrow +\infty$ справедливо асимптотическое равенство

$$h(y) = z_{**} + \sum_{k=1}^{\infty} \frac{z_k}{y^k}, \quad z_{**} \in (0, z_*), \quad (4)$$

сохраняющее силу при дифференцировании по y любое число раз.

Наглядное представление о свойствах функции $z = h(y)$ даёт её график, показанный на рис. 2. Что же касается поверхности (3), то в силу условий 1, 2 она имеет вид, представленный на рис. 3.

Действительно, обозначим через $y = y_1(z)$ и $y = y_2(z)$ корни уравнений $h(y) = z$ при $z \in (-\infty, z_*]$ и $h(y) = z$ при $z \in (z_{**}, z_*]$ из промежутков $(-\infty, y_*]$ и $[y_*, +\infty)$ соответственно (существование этих корней вытекает из условия 2). Далее, опираясь на условие 1, заключаем, что поверхность Γ разбивается на две части — устойчивую Γ_- (на которой $h'(y) > 0$) и неустойчивую Γ_+ (в точках которой $h'(y) < 0$), разделённые линией срыва Γ_0 (где $h'(y) = 0$). Для указанных частей справедливы равенства

$$\begin{aligned} \Gamma_- &= \{(x, y) : y = \Phi_-(x), x \in \Omega_1\}, \quad \Gamma_+ = \{(x, y) : y = \Phi_+(x), x \in \Omega_2\}, \\ \Gamma_0 &= \{(x, y) : y = y_*, x \in I_1\}, \end{aligned} \quad (5)$$

$$\Phi_-(x) = y_1(z)|_{z=g(x)}, \quad \Phi_+(x) = y_2(z)|_{z=g(x)}. \quad (6)$$

Добавим, что в равенствах (5) через $\Omega_1 \subset \mathbb{R}^2$ обозначена внутренняя область, ограниченная простой замкнутой кривой $l_1 = \{x \in \mathbb{R}^2 : g(x) = z_*\}$, а $\Omega_2 \subset \mathbb{R}^2$ представляет собой кольцевую область, ограниченную кривой l_1 и простой замкнутой кривой $l_2 = \{x \in \mathbb{R}^2 : g(x) = z_{**}\} \subset \Omega_1$. Заметим также, что поскольку $\lim_{x \in \Omega_2} \Phi_+(x) = +\infty$ при $x \in \Omega_2$, $x \rightarrow l_2$, то поверхность Γ представляет собой «кувшин с бесконечно длинным горлом» (см. рис. 3).

Теперь перейдём к ограничениям, касающимся вектор-функции $f_2(x)$ из выражения (1). В связи с этим рассмотрим вспомогательную систему

$$\dot{x} = f_2(x). \quad (7)$$

Условие 3. Предполагаем, что все траектории системы (7) с начальными условиями, принадлежащими кривой l_1 , за конечное время попадают на l_2 (рис. 4). Считаем также, что упомянутые траектории не имеют контактов с кривой l_2 , т. е.

$$(\text{grad } g(x), f_2(x)) < 0 \quad \text{при } x \in l_2, \quad (8)$$

где $(*, *)$ — евклидово скалярное произведение.

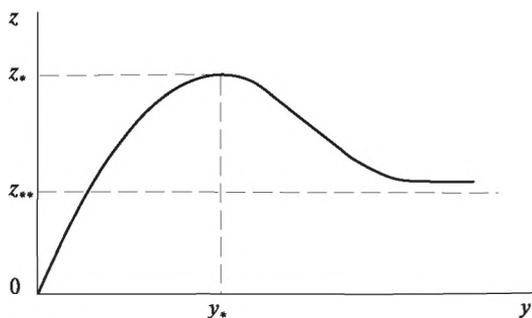


Рис. 2. График функции $z = h(y)$

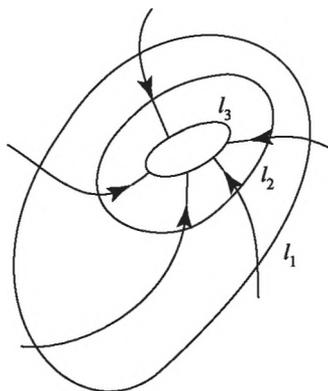


Рис. 4. Траектории вспомогательной системы (7)

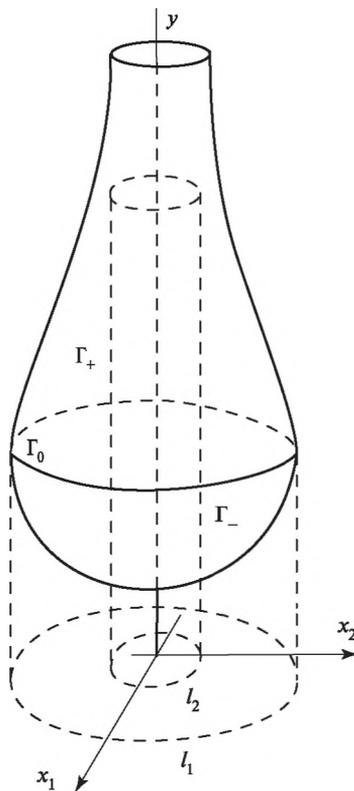


Рис. 3. Поверхность $\Gamma = \{(x, y) : g(x) - h(y) = 0\}$

Заключительная серия ограничений связана с системой

$$\dot{x} = f_1(x, \mu), \quad x \in \Omega_1 \cup l_1. \quad (9)$$

Однако прежде чем сформулировать соответствующие условия, введём в рассмотрение новую кривую l_3 , расположенную в области, ограниченной кривой l_2 (см. рис. 4). В связи с этим предпримем некоторые дополнительные построения.

Фиксируем произвольно точку $x = x_0$ лежащую в некоторой достаточно малой окрестности кривой l_1 , и обозначим через $x(t, x_0)$, $x(0, x_0) = x_0$ решение задачи Коши $\dot{x} = f_2(x)$, $x|_{t=0} = x_0$. Далее, введём в рассмотрение функцию

$$a(t, x_0) = \int_0^t [g(x(\theta, x_0)) - z_{**}] d\theta \quad (10)$$

и заметим, что она обладает свойствами:

$$\begin{aligned} a(0, x_0) = 0, \quad \frac{\partial a}{\partial t}(t, x_0) = g(x(t, x_0)) - z_{**} > 0 \text{ при } 0 \leq t < t_*, \\ \frac{\partial a}{\partial t}(t, x_0) < 0 \text{ при } t > t_*, \quad \lim_{t \rightarrow +\infty} a(t, x_0) = -\infty, \end{aligned} \quad (11)$$

где $t_* = t_*(x_0)$ — корень уравнения

$$g(x(t, x_0)) - z_{**} = 0. \quad (12)$$

Действительно, будем считать точку x_0 настолько близкой к кривой l_1 , что $g(x_0) - z_{**} > 0$, а для траектории $x = x(t, x_0)$ системы (7) сохраняется условие 3. Тогда эта траектория с течением времени ровно один раз пересечёт кривую l_2 и согласно неравенству (8) при последующем увеличении t останется в области $\{x \in \mathbb{R}^2 : g(x) - z_{**} < 0\}$, не приближаясь к её границе. А отсюда, в свою очередь, заключаем, что, во-первых, уравнение (12) имеет единственный корень $t = t_*(x_0) > 0$ с требуемыми свойствами (см. выражение (11)); во-вторых, при всех достаточно больших t подынтегральное выражение в уравнении (10) отрицательно и отделено от нуля. Следовательно, при $t \rightarrow +\infty$ заведомо выполняется и фигурирующее в (11) предельное равенство.

Опираясь на свойства (11), нетрудно увидеть, что уравнение $a(t, x_0) = 0$ имеет на полуоси $t > 0$ единственное решение $t = t_{**}(x_0)$. Учитывая это обстоятельство, введём в рассмотрение двумерное отображение

$$\Pi_0 : x_0 \rightarrow x(t, x_0)|_{t=t_{**}(x_0)}. \quad (13)$$

Несложная проверка показывает, что оно представляет собой диффеоморфизм, определённый в некоторой достаточно малой окрестности кривой l_1 .

Это значит, что $l_3 = \Pi_0(l_1)$ — простая замкнутая кривая класса C^∞ (см. рис. 4). Данная кривая и является искомой.

Возвращаясь к системе (9), предположим, что выполнен следующий блок ограничений.

Условие 4. Фазовый портрет системы (9) при $\mu = 0$ имеет вид, показанный на рис. 5: в области Ω_1 существуют единственное состояние равновесия $x = \bar{x}$, являющееся экспоненциально неустойчивым узлом или фокусом, и единственный окружающий его полуустойчивый цикл

$$L_0 : x = x_0(\varphi), \quad \frac{d\varphi}{dt} = \omega_0, \quad x_0(\varphi + 2\pi) \equiv x_0(\varphi), \quad \omega_0 > 0 \quad (14)$$

типа простой седло-узел. Предполагаем также, что все траектории из кольцевой области, ограничиваемой кривыми l_1 и L_0 , за конечное время попадают на кривую l_1 , пересекая её общим образом, т. е.

$$(\text{grad } g(x), f_1(x, 0)) > 0 \forall x \in l_1. \quad (15)$$

Наконец, считаем, что кривая l_3 лежит в области, ограничиваемой циклом L_0 , но не содержит и не окружает точку $x = \bar{x}$ (см. рис. 5).

Для пояснения фигурирующего в условии 4 требования простоты седло-узлов цикла (14) фиксируем произвольно точку $M_0 \in L_0$ и обозначим через Σ достаточно малый отрезок нормали к кривой L_0 в этой точке (см. рис. 5). Параметром s на секущей Σ будем считать расстояние от произвольной точки $M \in \Sigma$ до M_0 , взятое со знаком «+», если M лежит на внешней части нормали и со знаком «-» — в противном случае. Тогда, очевидно, что на множестве $\{(s, \mu) : |s| \leq s_0, |\mu| \leq \mu_0\}$, где $s_0, \mu_0 > 0$ подходящим образом малы, определено отображение последования Пуанкаре

$$s \rightarrow R(s, \mu), \quad R \in C^\infty \left[[-s_0, s_0] \times [-\mu_0, \mu_0] \right] \quad (16)$$

по траекториям системы (9), причём для $R(s, 0)$ в точке $s=0$ справедливо тейлоровское разложение вида

$$R(s, 0) = s + d_0 s^2 + O(s^3). \quad (17)$$

Простота седло-узлов цикла L_0 означает, что $d_0 > 0$.

Условие 5. Считаем, что

$$\alpha_0 = R'_\mu(0, 0) > 0, \quad (18)$$

где $R(s, \mu)$ — функция из выражения (16).

Из свойств (17), (18) и неравенства $d_0 > 0$ следует, что при $\mu \in [-\mu_0, 0)$ отображение (16) имеет две неподвижные точки $s_\pm(\mu) = \pm \sqrt{-\alpha_0 \mu / d_0} + O(\mu)$, которым в системе (9) соответствуют два цикла — устойчивый $L_-(\mu)$ и неустойчивый $L_+(\mu)$, $L_\pm(0) = L_0$. Таким образом, в этом случае все траектории рассматриваемой системы с начальными условиями, лежащими на кривой l_3 , при $t \rightarrow +\infty$ стремятся к циклу $L_-(\mu)$. Если же $\mu \in (0, \mu_0]$, то циклы в системе (9) отсутствуют и любая её траектория с начальным условием из l_3 сначала асимптотически долгое время (порядка $1/\sqrt{|\mu|}$) вращается в окрестности исчезнувшего цикла L_0 , а затем попадает на кривую l_1 .

При сформулированных ограничениях согласно результатам монографии [Мищенко и др., 1995] любая траектория

$$(x(t, \varepsilon, \mu), y(t, \varepsilon, \mu)) : x(0, \varepsilon, \mu) = x_0, \quad y(0, \varepsilon, \mu) = y_0 \quad (19)$$

системы (1) с независимыми от ε, μ начальными условиями x_0, y_0 на качественном уровне ведёт себя следующим образом.

Пусть точка (x_0, y_0) расположена в некоторой достаточно малой окрестности устойчивого многообразия медленных движений Γ_- и x_0 принадлежит

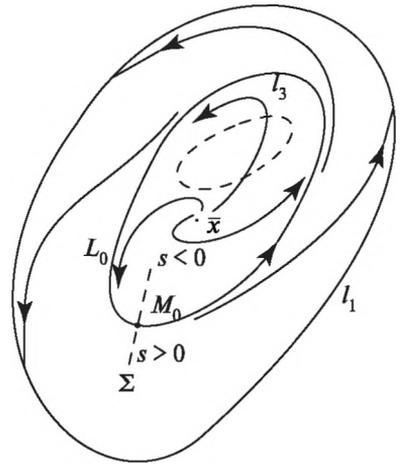


Рис. 5. Фазовый портрет системы (9) при $\mu = 0$

кольцевой области, ограничиваемой кривыми L_0 и l_1 . Тогда сначала за асимптотически малое время (порядка $\varepsilon \ln(1/\varepsilon)$) происходит «падение» фазовой точки (19) на поверхность Γ_- примерно по прямой $x = x_0$, а затем движение продолжается в асимптотически малой по ε окрестности кривой $(x_1(t, \mu), y_1(t, \mu))$, где $y_1(t, \mu) = \Phi_-(x_1(t, \mu))$, $\Phi_-(x)$ — функция из выражения (6), $x_1(t, \mu)$ — решение системы (9) с начальным условием $x_1(0, \mu) = x_0$. Далее, в силу условия 4 найдётся такой первый момент времени $t_1 = t_1(\mu) > 0$, что $x_1(t_1, \mu) \in l_1$. Отсюда вытекает [Мищенко и др., 1995], что при значениях t , асимптотически близких к t_1 , происходит «срыв» с многообразия Γ_- и начинается так называемый промежуток быстрого движения. Упомянутый промежуток, в свою очередь, можно разбить на три фрагмента: участок взлёта, поворота и падения.

На участке взлёта в силу вытекающего из условий 1, 2 неравенства

$$g(x) - h(y) > 0 \quad \forall x \in l_1, \quad \forall y > y_*$$

переменная y возрастает до асимптотически больших значений. Точнее говоря, фазовая точка (19) асимптотически быстро движется в асимптотически малой окрестности отрезка

$$\{(x, y) : x = x_1(\mu), y_* \leq y \leq \varepsilon^{-\delta}\}, \quad (20)$$

где $x(\mu) = x_1(t_1(\mu), \mu) \in l_1$, $\delta = (p_1 + 1)/(p_2 + 2) < 1$. Характерная особенность этого движения состоит в том, что любая часть отрезка (20), отвечающая промежутку $\bar{y} \leq y \leq \bar{y}$, где $y_* < \bar{y} < \varepsilon^{-\delta}$, проходится за время порядка ε . Величины же $y = \varepsilon^{-\delta}$ компонента $y(t, \varepsilon, \mu)$ достигает за время порядка $\varepsilon^{1-\delta}$.

Участок поворота соответствует значениям t , при которых фазовая точка (19) находится в полупространстве $\{(x, y) : y \geq \varepsilon^{-\delta}\}$. Для качественного описания этого участка выполним в системе (1) замены $y = u\varepsilon^{-\beta}$, $t - t_1(\mu) = \theta\varepsilon^{1-\beta}$, где $\beta = (p_1 + 1)/(p_2 + 1) < 1$ ($\delta < \beta < 1$) и воспользуемся разложением (4). В итоге после отбрасывания асимптотически малых по ε слагаемых приходим к задаче Коши

$$\frac{dx}{d\theta} = u^{p_2} f_2(x), \quad \frac{du}{d\theta} = g(x) - z_{**}, \quad x|_{\theta=0} = x_1(\mu), \quad u|_{\theta=0} = 0. \quad (21)$$

Полагая, что в выражении (21) $d\tau = u^{p_2} d\theta$, убеждаемся, что на участке поворота движение фазовой точки $(x(\tau, \varepsilon, \mu), u(\tau, \varepsilon, \mu))$, получающейся из соотношения (19) после сделанных замен, происходит в асимптотически малой окрестности кривой

$$\{(x, u) : x = x(\tau, \mu), u = u(\tau, \mu), \tau \geq 0\} \cap \{(x, u) : u \geq 0\}. \quad (22)$$

Здесь

$$\left. \begin{aligned} x(\tau, \mu) &= x(t, x_0) \Big|_{t=\tau, x_0=x_1(\mu)}, \\ u(\tau, \mu) &= \left((p_2 + 1) a(t, x_0) \right)^{1/(p_2 + 1)} \Big|_{t=\tau, x_0=x_1(\mu)}, \end{aligned} \right\} \quad (23)$$

а функции $x(t, x_0)$, $a(t, x_0)$ введены выше (см. (10)–(13)).

Обратим внимание, что в силу равенств (23) и известных свойств функции $a(t, x_0)$, фигурирующее в соотношении (22) априорное условие $u \geq 0$ выполняется только на отрезке $0 \leq \tau \leq \bar{\tau}(\mu)$, где $\bar{\tau}(\mu) = t_{**}(x_0) \Big|_{x_0=x_1(\mu)}$, а $t_{**}(x_0)$ — момент времени из выражения (13). Точнее говоря, компонента $u(\tau, \mu)$ обладает свойствами

$$u(\tau, \mu) > 0 \quad \text{при} \quad 0 < \tau < \bar{\tau}(\mu), \quad u(0, \mu) = u(\bar{\tau}(\mu), \mu) = 0. \quad (24)$$

Такое поведение $u(\tau, \mu)$ означает, что на участке поворота компонента $y(t, \varepsilon, \mu)$ сначала возрастает от $\varepsilon^{-\delta}$ до значений порядка $\varepsilon^{-\beta}$, а затем снова падает до прежнего уровня $y = \varepsilon^{-\delta}$. Добавим, что время прохождения данного участка равномерно по μ имеет порядок ε^{β} .

После поворота наступает так называемый участок падения, который вполне аналогичен описанному ранее участку взлёта. На этом участке траектория (19) асимптотически близка к отрезку

$$\{(x, y) : x = x_2(\mu), \Phi_-(x_2(\mu)) \leq y \leq \varepsilon^{-\delta}\}, \quad (25)$$

где $x_2(\mu) = x(\tau, \mu)|_{\tau=\bar{\tau}(\mu)}$. Заметим, далее, что в силу определения функций $x(\tau, \mu)$, $\bar{\tau}(\mu)$ из соотношений (23), (24) имеем $x_2(\mu) = \Pi_0(x_1(\mu))$, где Π_0 — оператор (13). Тем самым, автоматически $x_2(\mu) \in I_3$. Так как

$$g(x) - h(y) < 0 \forall x \in I_3, \forall y > \Phi_-(x),$$

то при движении вдоль отрезка (25) компонента $y(t, \varepsilon, \mu)$ убывает (отсюда и название — участок падения). Как и в случае участка взлёта, длительность данного участка имеет порядок $\varepsilon^{1-\delta}$.

За падением следует очередной участок медленного движения. На нём фазовая точка (19) находится в асимптотически малой окрестности кривой $(x_2(t, \mu), y_2(t, \mu))$, где $y_2(t, \mu) = \Phi_-(x_2(t, \mu))$, а $x_2(t, \mu)$ — решение задачи Коши для системы (9) с начальным условием $x_2|_{t=t_1(\mu)} = x_2(\mu)$. В зависимости от знака μ здесь возможны следующие два сценария.

Предположим, что параметр μ фиксирован и положителен. Тогда, как уже было сказано выше, система (9) не имеет циклов, а любое её решение с начальным условием на кривой I_3 за конечное, хотя и достаточно большое (порядка $1/\sqrt{\mu}$), время попадает на I_1 . Последнее означает существование такого момента $t = t_2 > t_1$, что при $t = t_2$ траектория (19) срывается с Γ_- . Затем идёт очередной участок быстрого движения и т. д. Ясно также, что в рассматриваемом случае описанный процесс смены участков быстрых и медленных движений продолжается до бесконечности, т. е. в системе (1) реализуются незатухающие неклассические релаксационные колебания.

Пусть теперь параметр μ фиксирован и отрицателен. Тогда, как следует из результатов статьи [Аносов, 1960], циклам $L_{\pm}(\mu)$ системы (9) в исходной системе (1) при всех достаточно малых $\varepsilon > 0$ соответствуют устойчивый и неустойчивый циклы $\tilde{L}_-(\varepsilon, \mu)$ и $\tilde{L}_+(\varepsilon, \mu)$, причём

$$\tilde{L}_{\pm}(0, \mu) = \{(x, y) : y = \Phi_{\pm}(x), x \in L_{\pm}(\mu)\}.$$

Отметим далее, что в силу условия 4 интересующая нас траектория (19) при $t > t_1(\mu)$ заведомо попадает в область притяжения цикла $\tilde{L}_-(\varepsilon, \mu)$. Таким образом, в данном случае релаксационные автоколебания в системе (1) отсутствуют.

Из приведённых качественных соображений вытекает существование критического значения $\mu_*(\varepsilon)$, $\mu_*(0) = 0$ параметра μ , при котором в системе (1) происходит переход от гладких автоколебаний — устойчивого цикла $\tilde{L}_-(\varepsilon, \mu)$, к релаксационным. Точнее говоря, случай $\mu = \mu_*(\varepsilon)$ соответствует седло-узловой бифуркации, приводящей к слиянию и исчезновению циклов $\tilde{L}_{\pm}(\varepsilon, \mu)$. Что же касается интересующей нас катастрофы голубого неба, то она наблюдается в системе (1) при

$$\mu = \mu_*(\varepsilon) + \nu, \quad 0 < \nu \ll 1 \quad (26)$$

и при некотором дополнительном условии, о котором будет сказано ниже.

1.2. Основные конструкции и утверждения

В этом разделе опишем математические конструкции, позволяющие сформулировать основные бифуркационные теоремы. Доказательства этих утверждений с минимальными изменениями могут быть заимствованы из статей [Глызин и др., 2008, 2015a] и поэтому здесь не приводятся.

Сформулируем сначала утверждение о том, что фигурирующее в выражении (26) критическое значение $\mu_*(\varepsilon)$ параметра μ действительно существует.

Лемма 1. *По любому натуральному k можно указать такое достаточно малое $\varepsilon_k > 0$, что на отрезке $0 \leq \varepsilon \leq \varepsilon_k$ существует единственная C^k -гладкая по $\sqrt{\varepsilon}$ функция $\mu_*(\varepsilon)$, $\mu_*(0) = 0$, обладающая следующим свойством: при $\forall \varepsilon \in (0, \varepsilon_0]$ и при $\mu = \mu_*(\varepsilon)$ система (1) имеет полуустойчивый цикл*

$$\tilde{L}_*(\varepsilon): x = x_*(\varphi, \varepsilon), \quad y = y_*(\varphi, \varepsilon), \quad \frac{d\varphi}{dt} = \omega_*(\varepsilon) \quad (27)$$

типа простой седло-узел. Здесь C^k -гладкие по $(\varphi, \sqrt{\varepsilon})$ функции $x_*(\varphi, \varepsilon)$, $y_*(\varphi, \varepsilon)$ и C^k -гладко зависящая от $\sqrt{\varepsilon}$ частота $\omega_*(\varepsilon)$ удовлетворяют равенствам:

$$\begin{aligned} x_*(\varphi + 2\pi, \varepsilon) &\equiv x_*(\varphi, \varepsilon), \quad y_*(\varphi + 2\pi, \varepsilon) \equiv y_*(\varphi, \varepsilon), \\ x_*(\varphi, 0) &= x_0(\varphi), \quad y_*(\varphi, 0) = \Phi_-(x_0(\varphi)), \quad \omega_*(0) = \omega_0, \end{aligned}$$

где вектор-функция $x_0(\varphi)$ и частота ω_0 заимствованы из выражения (14).

В дополнение отметим, что в силу условий 1–5 каждая траектория системы (1) при $\mu = \mu_*(\varepsilon)$, принадлежащая неустойчивому многообразию $W_{loc}^u(\tilde{L}_*(\varepsilon))$ цикла (27), при $t \rightarrow +\infty$ снова стремится к $\tilde{L}_*(\varepsilon)$, попадая в узловую область этого цикла. Таким образом $\mu_*(\varepsilon)$ является искомым критическим значением, соответствующим катастрофе голубого неба.

На следующем этапе введём в рассмотрение инвариантное многообразие медленных движений в окрестности цикла (14)

$$\{(x, y): y = H(x, \varepsilon, \mu), x \in \mathcal{U}\}, \quad (28)$$

где \mathcal{U} — достаточно малая окрестность цикла (14), функция $H(x, \varepsilon, \mu)$, $H(x, 0, \mu) = \Phi_-(x)$ является C^k -гладкой по совокупности переменных $(x, \sqrt{\varepsilon}, \mu)$. На многообразии (28) рассмотрим двумерную систему

$$\dot{x} = f(x, H(x, \varepsilon, \mu), \mu), \quad x \in \mathcal{U} \quad (29)$$

и учтём в ней равенства (26). Полученную при этом систему

$$\dot{x} = F_*(x, \varepsilon, \nu) \quad (30)$$

преобразуем по возможности к наиболее простому виду. Для этого введём в области \mathcal{U} радиальную и циклическую координаты $(s, \varphi): |s| \leq s_0, s_0 = \text{const} > 0, 0 \leq \varphi \leq 2\pi \pmod{2\pi}$, связанные с полуустойчивым циклом. Точнее говоря, в качестве φ возьмём угловую координату из соотношения (27), а через s обозначим параметр на отрезке нормали $\Sigma(\varphi, \varepsilon)$ к кривой $L_*(\varepsilon)$, проведённой через точку $x_*(\varphi, \varepsilon)$ (способ определения этого параметра описан выше). Далее, из результатов статьи [Глызин и др., 2007] вытекает существование такого достаточно малого $\rho_0 > 0$ и таких C^k -гладких по совокупности переменных $(\rho, \theta, \sqrt{\varepsilon}) \in [-\rho_0, \rho_0] \times [0, 2\pi] \times [0, \varepsilon_k]$ функций $h_j(\rho, \theta, \varepsilon)$, $h_j(\rho, \theta + 2\pi, \varepsilon) \equiv h_j(\rho, \theta, \varepsilon)$, $j = 1, 2$, что после перехода в системе (30) при $\nu = 0$

к координатам (s, φ) и замены $s = \rho + \rho^2 h_1(\rho, \theta, \varepsilon)$, $\varphi = \theta + \rho h_2(\rho, \theta, \varepsilon)$ она принимает вид

$$\dot{\rho} = d_1(\varepsilon)\rho^2 + d_2(\varepsilon)\rho^3, \quad \dot{\theta} = \omega_*(\varepsilon). \quad (31)$$

Здесь $\omega_*(\varepsilon)$ — частота из выражения (27); $d_j(\varepsilon)$, $j=1, 2$ — некоторые функции класса C^k по переменной $\sqrt{\varepsilon}$, причём $d_1(0) = d_0/2\pi > 0$, где d_0 — постоянная из соотношения (17).

Предположим теперь, что в системе (30) $\nu > 0$. Тогда указанные выше замены преобразуют её к виду

$$\dot{\rho} = d_1(\varepsilon)\rho^2 + d_2(\varepsilon)\rho^3 + \nu\Delta_1(\rho, \theta, \varepsilon, \nu), \quad \dot{\theta} = \omega_*(\varepsilon) + \nu\Delta_2(\rho, \theta, \varepsilon, \nu), \quad (32)$$

где $\Delta_j(\rho, \theta, \varepsilon, \nu)$, $j=1, 2$ — некоторые C^k -гладкие по $(\rho, \theta, \sqrt{\varepsilon}, \nu)$ и 2π -периодические по θ функции. В книге [Шильников и др., 2004] показано, что существует замена переменных

$$\begin{aligned} \rho &= r + \nu h_{1,1}(\psi, \varepsilon) + \nu r h_{1,2}(\psi, \varepsilon), \\ \theta &= \psi + \nu h_{2,1}(\psi, \varepsilon) + \nu r h_{2,2}(\psi, \varepsilon), \quad \text{где } h_{j,s}(\psi, \varepsilon), \quad h_{j,s}(\psi + 2\pi, \varepsilon) \equiv h_{j,s}(\psi, \varepsilon), \\ j, s &= 1, 2 \text{ — некоторые } C^k\text{-гладкие по } \psi, \sqrt{\varepsilon} \text{ функции, приводящая систему (32) к виду} \end{aligned}$$

$$\begin{cases} \dot{r} = d_1(\varepsilon)r^2 + d_2(\varepsilon)r^3 + \alpha_1(\varepsilon)\nu + \alpha_2(\varepsilon)\nu r + \nu\Delta_1(r, \psi, \varepsilon, \nu), \\ \dot{\psi} = \omega_*(\varepsilon) + \beta_1(\varepsilon)\nu + \beta_2(\varepsilon)\nu r + \nu\Delta_2(r, \psi, \varepsilon, \nu). \end{cases} \quad (33)$$

Здесь $\alpha_j(\varepsilon)$, $\beta_j(\varepsilon)$, $j=1, 2$ — скалярные функции; C^k -гладко зависящие от $\sqrt{\varepsilon}$, а остатки Δ_j , $j=1, 2$ в соотношении (33) удовлетворяют тем же общим требованиям периодичности и гладкости, что и аналогичные им остатки из выражения (32).

Последующий анализ базируется на некоторых дополнительных геометрических построениях. В связи с этим введём в рассмотрение две двумерные цилиндрические поверхности S_{\pm} , задающиеся в переменных (r, ψ, ν) , $\nu = y - H(x, \varepsilon, \mu)$ равенствами

$$S_{\pm} = \{(r, \psi, \nu) : r = \pm r_0, 0 \leq \psi \leq 2\pi \pmod{2\pi}, |\nu| \leq \nu_0\}, \quad (34)$$

где постоянные r_0 , $\nu_0 > 0$ фиксированы и достаточно малы. Отметим далее, что при условии (26) поверхности (34) не имеют контактов с траекториями системы (1), поскольку положительны производные $\dot{r}|_{r=\pm r_0}$ в силу этой системы (данный факт — следствие положительности коэффициентов $d_1(\varepsilon)$, $\alpha_1(\varepsilon)$ нормальной формы (33)). Кроме того, из описанного выше качественного характера поведения решений (19) вытекает, что, во-первых, в данном случае корректно определён оператор соответствия $\Pi_{+,-}(\varepsilon, \nu) : S_+ \rightarrow S_-$ по траекториям системы (1); во-вторых, при $\varepsilon \rightarrow 0$, $\nu \rightarrow 0$ этот оператор допускает конечный предел

$$\Pi_{+,-}(0, 0) : (\psi, \nu) \rightarrow (\gamma(\psi), 0), \quad (35)$$

где $\gamma(\psi) \in C^\infty$ — некоторая периодическая с периодом 2π функция.

Применяя к системе (1) общие результаты монографии [Мищенко и др., 1995], касающиеся классических и неклассических релаксационных колебаний, приходим к следующему утверждению.

Лемма 2. При выполнении условий 1–5 и соотношения (26) имеет место предельное равенство

$$\lim_{\varepsilon \rightarrow 0, \nu \rightarrow 0} \Pi_{+,-}(\varepsilon, \nu) = \Pi_{+,-}(0, 0) \quad (37)$$

в метрике C^k по переменным $(\psi, \nu) \in [0, 2\pi] \times [-\nu_0, \nu_0]$ при любом фиксированном k .

Заключительный этап исследования поведения решений системы (1) при условии (26) связан с рассмотрением оператора соответствия $\Pi_{-,+}(\varepsilon, \nu)$ по её траекториям, действующего из S_- в S_+ . Проблема существования оператора $\Pi_{-,+}(\varepsilon, \nu)$ сводится к аналогичному вопросу для оператора соответствия $\Pi_{-,+}^*(\varepsilon, \nu): L_{(-)} \rightarrow L_{(+)}$ по траекториям системы (33). Для этого оператора справедливо следующее утверждение.

Лемма 3. По любому натуральному k можно указать такие достаточно малые постоянные $\varepsilon_k > 0$, $\nu_k > 0$, что при всех $0 < \nu \leq \nu_k$, $0 \leq \varepsilon \leq \varepsilon_k$ оператор $\Pi_{-,+}^*(\varepsilon, \nu)$ допускает представление

$$\Pi_{-,+}^*(\varepsilon, \nu) : \psi \rightarrow \psi + \frac{c_*(\varepsilon, \nu)}{\sqrt{\nu}} + \sqrt{\nu} \cdot \Psi(\psi, \varepsilon, \nu) \pmod{2\pi}, \quad (37)$$

где функция $c_*(\varepsilon, \nu)$, $c_*(0, 0) > 0$ непрерывна по совокупности переменных (вплоть до значения $\nu = 0$). Что же касается 2π -периодической по ψ функции $\Psi(\psi, \varepsilon, \nu)$, то она непрерывна и ограничена на множестве $(\psi, \varepsilon, \nu) \in [0, 2\pi] \times [0, \varepsilon_k] \times [0, \nu_k]$ вместе со своими производными $\partial^m \Psi / \partial \psi^m$, $m \leq k$.

Обратимся теперь к исходному оператору $\Pi_{-,+}(\varepsilon, \nu)$. Из проделанных выше построений вытекает следующее.

Лемма 4. Для любого натурального k найдутся такие достаточно малые $\varepsilon_k > 0$, $\nu_k > 0$, что при всех $0 < \nu \leq \nu_k$, $0 \leq \varepsilon \leq \varepsilon_k$ оператор $\Pi_{-,+}(\varepsilon, \nu)$ допускает представление

$$\begin{aligned} \Pi_{-,+}(\varepsilon, \nu) : \\ \psi \rightarrow \psi + \frac{c_*(\varepsilon, \nu)}{\sqrt{\nu}} + \Psi_1(\psi, \nu, \varepsilon, \nu) \pmod{2\pi}, \\ \nu \rightarrow \Psi_2(\psi, \nu, \varepsilon, \nu) \exp \left[-\frac{c_{**}}{\varepsilon \sqrt{\nu}} \right]. \end{aligned} \quad (38)$$

Здесь функция $c_*(\varepsilon, \nu)$ обладает теми же свойствами, что и аналогичная функция из выражения (37), $c_{**} = \text{const} > 0$, а 2π -периодические по ψ функции Ψ_j , $j = 1, 2$ и их всевозможные частные производные по ψ, ν до порядка k непрерывны по $(\psi, \nu, \varepsilon, \nu) \in [0, 2\pi] \times [-\nu_0, \nu_0] \times [0, \varepsilon_k] \times [0, \nu_k]$. Кроме того, выполняется равенство $\Psi_1(\psi, \nu, 0, 0) \equiv 0$.

Проделанные построения позволяют определить оператор последования Пуанкаре

$$\Pi(\varepsilon, \nu) = \Pi_{-,+}(\varepsilon, \nu) \circ \Pi_{+,-}(\varepsilon, \nu) : S_+ \rightarrow S_+ \quad (39)$$

по траекториям системы (1) при условии (26). Из соотношений (37), (38) вытекает следующее основное утверждение.

Теорема 1. Для любого натурального k найдутся такие достаточно малые $\varepsilon_k > 0$, $\nu_k > 0$, что при $0 \leq \varepsilon \leq \varepsilon_k$, $0 < \nu \leq \nu_k$ оператор (39) имеет вид

$$\begin{aligned} \Pi(\varepsilon, \nu) : \\ \psi \rightarrow \frac{c_*(\varepsilon, \nu)}{\sqrt{\nu}} + \gamma(\psi) + \Lambda_1(\psi, \nu, \varepsilon, \nu) \pmod{2\pi}, \\ \nu \rightarrow \Lambda_2(\psi, \nu, \varepsilon, \nu) \exp \left[-\frac{c_{**}}{\varepsilon \sqrt{\nu}} \right]. \end{aligned} \quad (40)$$

Здесь $\gamma(\psi)$ и $c_*(\varepsilon, \nu)$ — функции из соотношений (35) и (38) соответственно; $c_{**} = \text{const} > 0$; 2π -периодические по ψ функции Λ_j , $j = 1, 2$ и их производные по (ψ, ν) до порядка k включительно непрерывны по

$(\psi, \nu, \varepsilon, \nu) \in [0, 2\pi] \times [-\nu_0, \nu_0] \times [0, \varepsilon_k] \times [0, \nu_k]$. Кроме того, выполняется равенство $\Lambda_1(\psi, \nu, 0, 0) \equiv 0$.

Опираясь на представление (40), нетрудно увидеть, что $\Pi(\varepsilon, \nu)S_{\mp} \subset S_{\mp}$. А это значит, что отображение (39) имеет максимальный аттрактор

$$A_{\max} = \bigcap_{n=0}^{\infty} \Pi^n(\varepsilon, \nu)S_{\mp}. \quad (41)$$

Ясно также, что в первом приближении за структуру множества (41) отвечает одномерное отображение

$$\Pi_{\varkappa} : \psi \rightarrow \bar{\psi} = \varkappa + \gamma(\psi), \quad (42)$$

где $\varkappa = c_*(\varepsilon, \nu)/\sqrt{\nu}$, получающееся из выражения (40) после отбрасывания асимптотически малых по ε, ν слагаемых. В дальнейшем будем рассматривать это отображение, считая \varkappa независимым параметром, пробегающим всю числовую ось \mathbb{R} .

При анализе аттрактора (41) ограничимся двумя основными случаями. Начнём с простейшего из них, а именно, предположим, что

$$|\gamma'(\psi)| < 1 \forall \psi \in [0, 2\pi]. \quad (43)$$

Тогда, как нетрудно увидеть, отображение (42) имеет единственную непрерывно зависящую от \varkappa неподвижную точку

$$\psi = \psi_0(\varkappa), \quad \psi_0(\varkappa + 2\pi) \equiv \psi_0(\varkappa) + 2\pi. \quad (44)$$

Что же касается исходного отображения (40), то при условии (43) оно очевидным образом является сжимающим. Тем самым, в рассматриваемом случае приходим к следующему утверждению.

Теорема 2 (о катастрофе голубого неба). Пусть выполнено неравенство (43). Тогда аттрактор (41) состоит из единственной экспоненциально устойчивой неподвижной точки $(\psi, \nu) = (\psi(\varepsilon, \nu), \nu(\varepsilon, \nu))$, для компонент которой при $\nu \rightarrow 0$ справедливы равномерные по ε асимптотические представления

$$\psi(\varepsilon, \nu) = \left((1 + o(1))\psi_0(\varkappa) \right) \Big|_{\varkappa = \frac{c_*(\varepsilon, \nu)}{\sqrt{\nu}}}, \quad \nu(\varepsilon, \nu) = O \left| \exp \left[-\frac{c_{**}}{\varepsilon\sqrt{\nu}} \right] \right|. \quad (45)$$

где $c_{**} > 0$ — постоянная из выражения (40); $\psi_0 = \psi_0(\varkappa)$ — функция (44).

Следует отметить, что теорема 2 гарантирует реализуемость в системе (1) при условии (26) интересующей нас катастрофы голубого неба. Действительно, неподвижной точке (45) оператора последования (39) отвечает устойчивый релаксационный цикл $\bar{L}(\varepsilon, \nu)$ этой системы, период и длина которого равномерно по ε стремятся к бесконечности при $\nu \rightarrow 0$. Кроме того, в силу первого равенства из соотношения (45) множество всех частичных пределов компоненты $\psi(\varepsilon, \nu)$ при $\nu \rightarrow 0$, взятое по модулю 2π , совпадает с отрезком $[0, 2\pi]$. Это значит, что верхний топологический предел кривой $\bar{L}(\varepsilon, \nu)$ при $\nu \rightarrow 0$ равен $W^u(\bar{L}_*(\varepsilon)) \cup \bar{L}_*(\varepsilon)$, где $\bar{L}_*(\varepsilon)$ — седло-узловой цикл (27), а W^u — его неустойчивое многообразие, продолженное по траекториям системы (1) при $\mu = \mu_*(\varepsilon)$ и состоящее из двоякоасимптотических к $\bar{L}_*(\varepsilon)$ решений.

Предположим теперь, что в условии 4 кривая l_3 по-прежнему лежит в области, ограничиваемой циклом L_0 , но окружает состояние равновесия $x = \bar{x}$ и не имеет контактов с траекториями системы (9) при $\mu = 0$. В этом случае фигурирующая в выражении (35) функция $\gamma(\psi)$ обладает свойствами

$$\gamma(\psi + 2\pi) \equiv \gamma(\psi) + 2\pi, \quad \gamma'(\psi) > 0 \quad \forall \psi \in [0, 2\pi]. \quad (46)$$

Что же касается всех приведённых выше конструкций и итоговой теоремы 1, то они в рассматриваемой ситуации остаются в силе. Более того, справедливо следующее утверждение.

Теорема 3. Пусть условие 4 модифицировано описанным выше образом. Тогда аттрактор (41) отображения (40) состоит из инвариантной кривой

$$\left\{ (\psi, \nu) : \nu = V(\psi, \varepsilon, \nu) \exp \left[-\frac{c_{**}}{\varepsilon \sqrt{\nu}} \right] \right\}. \quad (47)$$

Здесь $c_{**} > 0$, а 2π -периодическая по ψ функция $V(\psi, \varepsilon, \nu)$ обладает теми же свойствами непрерывности и гладкости, что и функция $\Lambda_2(\psi, 0, \varepsilon, \nu)$ из выражения (40).

Для доказательства этого утверждения используется известный принцип кольца (см. [Колесов, Розов, 2004; Шильников и др., 2004]). Из этого принципа вытекает существование в множестве S_+ единственной глобально экспоненциально устойчивой инвариантной кривой вида (47). Добавим ещё, что в исходной системе (1) при условии (26) данной кривой отвечает устойчивый двумерный инвариантный тор.

2. ЧИСЛЕННЫЙ АНАЛИЗ МОДЕЛЬНОЙ СИСТЕМЫ

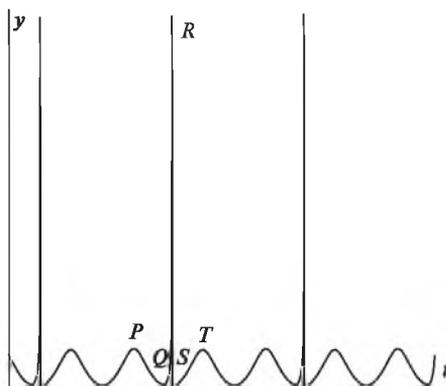
В качестве конкретного примера модели (1) возьмём систему, в которой

$$\begin{aligned} f_1(x, \mu) &= (f_{1,1}, f_{1,2}), \quad f_2(x) = (f_{2,1}, f_{2,2}), \\ f_{1,1} &= (x_1 - b_1) \frac{\rho}{\sqrt{1 + \rho^2}} - x_2, \quad f_{1,2} = x_2 \frac{\rho}{\sqrt{1 + \rho^2}} + x_1 - b_1, \\ \rho &= \left[(x_1 - b_1)^2 + x_2^2 - b_2^2 \right] + \mu, \quad f_{2,1} = -\alpha x_1, \quad f_{2,2} = -\alpha x_2, \\ g(x) &= \delta_1 x_1^2 + \delta_2 x_2^2, \quad h(y) = c_1 y \exp(-y) + c_2 (1 - \exp(-y)), \end{aligned} \quad (48)$$

а параметры заданы равенствами $b_1 = 1$, $b_2 = 2$, $c_1 = 28$, $c_2 = 1$, $\alpha = 0,5$, $\delta_1 = \delta_2 = 1$, $\varepsilon = 0,01$, $\mu = 0,0015$. Выбор параметров p_1 , p_2 позволяет менять порядок максимальных значений решения системы (1). Например, если положить $p_1 = 1/2$, $p_2 = 5$, то колебания компоненты y на устойчивом цикле системы (1) при значениях параметров будут иметь вид, показанный на рис. 6. В данном случае значения высоты P -зубцов являются величинами порядка $\varepsilon^{-1/4}$.

Как показал численный анализ, в этом случае аттрактором нашей системы является устойчивый релаксационный цикл. График зависимости от t компоненты y данного цикла представлен на рис. 6. Сравнивая его с графиком реальной ЭКГ (см. рис. 1), мы видим, что и здесь имеются характерные зубцы P , Q , R , S , T , о которых говорилось выше.

Следует отметить, что здоровое человеческое сердце, вообще говоря, не является периодическим осциллятором. Скорее всего, кардиосигнал квазипериодичен [Казаков, 2012] или даже хаотичен [Babloyantz, Destexhe, 1998]. Построенная нами теория позволяет без труда смоделировать квазипериодический вариант (пример приведён в статье [Глызин и др., 2015а] при $p_1 = 1/2$, $p_2 = 2$). Для моделирования хаотического кардиосигнала целесообразно рассмотреть многомерный вариант системы (1), в котором $x \in \mathbb{R}^m$, $m \geq 3$, а условия 1–5 надлежащим образом модифицированы. Точнее говоря, условия 1 и 2 здесь сохраняются практически дословно.


 Рис. 6. График компоненты $y(t)$ системы (1), (48)

Однако теперь мы предполагаем, что при любом $z > 0$ уравнение $g(x) = z$ определяет замкнутую $(m - 1)$ -мерную поверхность класса C^∞ , гомеоморфную сфере. Условия 3 и 4 в отличие от двух предыдущих нуждаются в более существенной переработке, а именно, мы считаем, что в области $\Omega_1 = \{x \in \mathbb{R}^m : g(x) < z_*\}$ система (9) при $\mu = 0$ по-прежнему имеет цикл L_0 типа простой седло-узла. Далее, предполагаем, что траектории этой системы с начальными условиями, лежащими на неустойчивом многообразии $W^u(L_0)$ данного цикла, с течением времени пересекают поверхность $l_1 = \{x \in \mathbb{R}^m : g(x) = z_*\}$, но не имеют контактов с ней. Таким образом, множество $W^u(L_0) \cap l_1$ представляет собой простую замкнутую кривую C_1 класса C^∞ .

Следующее ограничение касается поведения траекторий системы (9) при $\mu = 0$ с начальными условиями на кривой $C_2 = \Pi_0(C_1)$, где Π_0 — диффеоморфизм (13). Будем считать, что все эти траектории при $t \rightarrow +\infty$ стремятся к циклу L_0 , касаясь его двумерного экспоненциально устойчивого центрального многообразия $W^c(L_0)$.

Для того чтобы сформулировать аналог условия 5, введём в рассмотрение экспоненциально устойчивое двумерное инвариантное многообразие W_μ системы (9), являющееся продолжением по μ многообразия $W^c(L_0)$, о котором говорилось выше. Далее, рассмотрим двумерную систему, представляющую собой сужение (9) на W_μ , и определим для неё аналогичное (16) отображение последования Пуанкаре. Как и прежде, предполагаем, что в тейлоровском разложении (17) этого отображения коэффициент d_0 строго положителен и выполняется неравенство вида (18).

Из проделанных выше построений, сохраняющих силу и в многомерном случае, вытекает, что при $x \in \mathbb{R}^m$, $m \geq 3$ за аттракторы системы (1), (26) отвечает аналогичное (40) отображение

$$\begin{aligned} \psi &\rightarrow \frac{c_*(\varepsilon, \nu)}{\nu} + \gamma(\psi) + \Lambda_1(\psi, \nu_1, \nu_2, \varepsilon, \nu) \pmod{2\pi}, \\ \Pi(\varepsilon, \nu) : \nu_1 &\rightarrow \Lambda_2(\psi, \nu_1, \nu_2, \varepsilon, \nu) \exp\left[-\frac{c_{**}}{\varepsilon\sqrt{\nu}}\right], \\ \nu_2 &\rightarrow \Lambda_3(\psi, \nu_1, \nu_2, \varepsilon, \nu) \exp\left[-\frac{c_{***}}{\sqrt{\nu}}\right], \end{aligned} \quad (49)$$

определённое в некотором кольце

$$K = \{(\psi, v_1, v_2) : 0 \leq \psi \leq 2\pi \pmod{2\pi}, |v_1| \leq v_{0,1}, \|v_2\| \leq v_{0,2}\}. \quad (50)$$

Здесь $v_1 \in \mathbb{R}$, $v_2 \in \mathbb{R}^{m-1}$, $\|\cdot\|$ — евклидова норма в \mathbb{R}^{m-1} , постоянные $v_{0,1}, v_{0,2} > 0$ достаточно малы, функция $c_\varepsilon(\varepsilon, \nu)$ обладает теми же свойствами, что и аналогичная функция в выражении (40), $c_{**}, c_{***} = \text{const} > 0$. Что же касается функций Λ_j , $j = 1, 2, 3$, то они непрерывны по совокупности переменных (вплоть до значений $\varepsilon = 0, \nu = 0$) вместе со своими частными производными по (ψ, v_1, v_2) до порядка k включительно (k — любое наперёд заданное натуральное число). И наконец, следует добавить, что $\Lambda_1(\psi, v_1, 0, 0, 0) \equiv 0$.

Как и в случае $m = 2$, за аттракторы отображения (49) в кольце (50) отвечает в конечном итоге одномерное отображение окружности в себя вида (42). В свою очередь, фигурирующая в нём функция $\gamma(\psi) \in C^\infty$ допускает представление

$$\gamma(\psi) = n_0 \psi + \gamma_0(\psi), \quad \gamma_0(\psi + 2\pi) \equiv \gamma_0(\psi), \quad (51)$$

где $n_0 \in \mathbb{Z}$ — некоторый топологический инвариант.

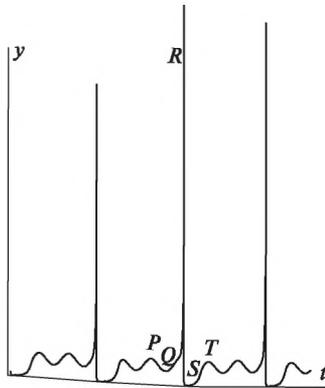


Рис. 7. График компоненты $y(t)$ системы (1), (53)

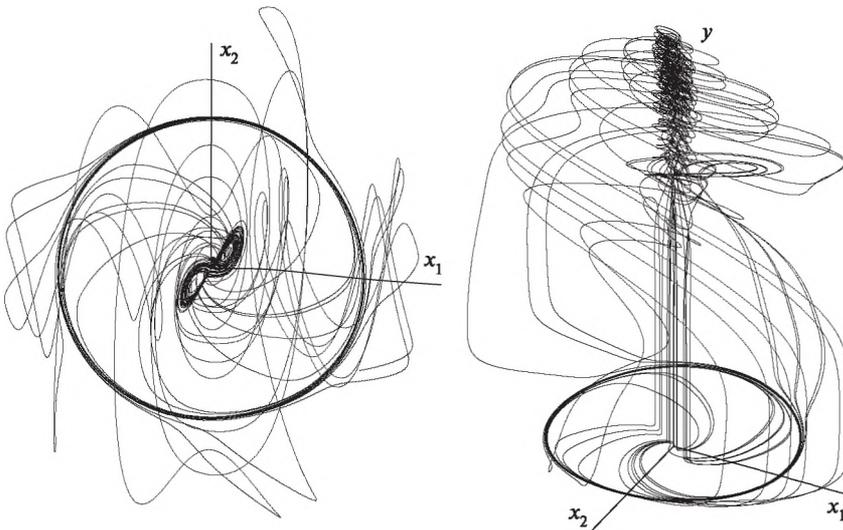


Рис. 8. Проекции хаотического аттрактора A системы (1), (53)

При $m = 2$ в формуле (51) возможны лишь случаи $n_0 = 0$ и $n_0 = 1$. В первом из них кривая l_3 не окружает особую точку $x = \bar{x}$ (см. рис. 5), а во втором окружает. Выше для каждого из этих случаев был получен свой результат (см. теоремы 2 и 3). Данные результаты сохраняются, естественно, и при $m \geq 3$. Однако уже при $m = 3$ возможна принципиально новая ситуация, когда

$$|n_0| \geq 2, \quad |n_0 + \gamma'_0(\psi)| > 1 \forall \psi \in [0, 2\pi]. \quad (52)$$

Из содержащихся в статье [Тураев, Шильников, 1995] построений следует, что при условиях (52) отображение (49) имеет в кольце (50) хаотический гиперболический аттрактор типа соленоида Смейла-Вильямса.

Конкретным примером многомерной модели (1) при $x \in \mathbb{R}^3$ служит система, в которой

$$\begin{aligned} f_1 &= \left(x_1 - b_1 - \frac{\rho}{\sqrt{1+\rho^2}} - x_2, x_2 \frac{\rho}{\sqrt{1+\rho^2}} + x_1 - b_1, -\chi x_3 \right), \\ \rho &= \left[(x_1 - b_1)^2 + x_2^2 - b_2^2 \right]^2 + \mu, \\ f_2 &= \alpha \left(10(x_2 - x_1), 28x_1 - x_2 - 40x_1x_3, -\frac{8x_3}{3} + 40x_1x_2 \right), \\ g(x) &= \delta_1 x_1^2 + \delta_2 x_2^2 + \delta_3 x_3^2, \quad h(y) = c_1 y \exp(-y) + c_2 (1 - \exp(-y)). \end{aligned} \quad (53)$$

Как показал численный эксперимент, при значениях параметров $b_1 = 0$, $b_2 = 2$, $c_1 = 18$, $c_2 = 1$, $\alpha = 0,5$, $\chi = 10$, $\varepsilon = 0,005$, $\mu = 0,0035$, $\delta_1 = 1$, $\delta_2 = 0,6$, $\delta_3 = 1,9$, $p_1 = 0,5$, $p_2 = 5$ система (1), (53) имеет хаотический аттрактор A с ляпуновскими показателями $\lambda_1 \approx 1,1$, $\lambda_2 = 0$, $\lambda_3 \approx -10,31$, $\lambda_4 \approx -2763$ и ляпуновской размерностью $d_L \approx 2,1$ (вычисляющейся по известной формуле Каплана-Йорке). График зависимости от t компоненты y на аттракторе A представлен на рис. 7. Отметим, что зубец R асимптотически высок и согласно нашей асимптотической теории имеет порядок $\varepsilon^{-\beta}$, где $\beta = (p_1 + 1)/(p_2 + 1)$. Учитывая, что $p_1 = 0,5$, $p_2 = 5$, получаем порядок максимума равный $\varepsilon^{-1/4}$. Для иллюстрации аттрактора A на рис. 8 приведены две проекции траекторий системы (1), (53).

ЛИТЕРАТУРА

- [Аносов, 1960] Аносов Д. В. О предельных циклах систем дифференциальных уравнений с малым параметром при старших производных // Матем. сб. 1960. Т. 50. № 3. С. 299–334.
- [Глызин и др., 2007] Глызин С. Д., Колесов А. Ю., Розов Н. Х. Разделение движений в окрестности полуустойчивого цикла // Дифференц. уравнения. 2007. Т. 43. № 5. С. 598–615.
- [Глызин и др., 2008] Глызин С. Д., Колесов А. Ю., Розов Н. Х. Катастрофа голубого неба в релаксационных системах с одной быстрой и двумя медленными переменными // Дифференц. уравнения. 2008. Т. 44. № 2. С. 158–171.
- [Глызин и др., 2015а] Глызин С. Д., Колесов А. Ю., Розов Н. Х. Катастрофа голубого неба в системах с неклассическими релаксационными колебаниями // Моделирование и анализ информ. систем. 2015. Т. 22. № 1. С. 38–64.
- [Глызин и др., 2015б] Глызин С. Д., Колесов А. Ю., Розов Н. Х. «Катастрофа голубого неба» в приложении к моделированию кардиоритмов // Журн. вычислит. математики и математич. физики. 2015. Т. 55. № 7. С. 1136–1155.

- [Казаков, 2012] *Казаков Д. В.* Квазипериодическая двухкомпонентная динамическая модель для синтеза кардиосигнала с использованием временных рядов и метода Рунге-Кутты четвертого порядка // Компьютерные исслед. и моделирование. 2012. Т. 4. № 1. С. 143–154.
- [Колесов, Розов, 2004] *Колесов А. Ю., Розов Н. Х.* Инвариантные торы нелинейных волновых уравнений. М.: Физматлит, 2004. 408 с.
- [Мищенко, Розов, 1975] *Мищенко Е. Ф., Розов Н. Х.* Дифференциальные уравнения с малым параметром и релаксационные колебания. М.: Наука, 1975. 248 с.
- [Мищенко и др., 1995] *Мищенко Е. Ф., Колесов Ю. С., Колесов А. Ю., Розов Н. Х.* Периодические движения и бифуркационные процессы в сингулярно возмущённых системах. М.: Физматлит, 1995. 336 с.
- [Стрыгин, Соболев, 1988] *Стрыгин В. В., Соболев В. А.* Разделение движений методом интегральных многообразий. М.: Наука, 1988. 256 с.
- [Тураев, Шильников, 1995] *Тураев Д. В., Шильников Л. П.* О катастрофах голубого неба // ДАН. 1995. Т. 342. № 5. С. 596–599.
- [Шильников и др., 2004] *Шильников Л. П., Шильников А. Л., Тураев Д. В., Чуа Л.* Методы качественной теории в нелинейной динамике. Ч. 1. Москва-Ижевск: Институт компьютерных исследований, 2004. 416 с.
- [Babloyantz, Destexhe, 1998] *Babloyantz A., Destexhe A.* Is the Normal Heart a Periodic Oscillator? // Biol. Cybern. 1998. V. 58. P. 203–211.
- [Janson et al., 1998] *Janson N. B., Pavlov A. N., Anishenko V. S.* One method for restoring inhomogeneous attractors // J. Bifurcation and Chaos. 1998. V. 8. No. 4. P. 825–833.
- [McSharry et al., 2003] *McSharry P. E., Clifford G. D., Tarassenko L., Smith L. A.* A Dynamical Model for Generating Synthetic Electrocardiogram Signals // IEEE Trans. Biomed. Eng. 2003. V. 50. No. 3. P. 289–294.
- [Podziemski, Zebrowski, 2013] *Podziemski P., Zebrowski J. J.* A simple model of the right atrium of the human heart with the sinoatrial and atrioventricular nodes included // J. Clin. Monit. Comput. 2013. V. 27. P. 481–498.
- [dos Santos et al., 2004] *dos Santos A. M., Lopes S. R., Viana R. L.* Rhythm synchronization and chaotic modulation of coupled Van der Pol oscillators in a model for the heartbeat // Physica A. 2004. V. 338. P. 335–355.
- [Shilnikov et al., 2003] *Shilnikov A., Shilnikov L., Turaev D.* Blue sky catastrophe in singularly-perturbed systems: Preprint WIAS. No. 841. Berlin, 2003.
- [Shilnikov et al., 2005] *Shilnikov A., Shilnikov L., Turaev D.* Blue-sky catastrophe in singularly perturbed systems // Moscow Mathematical J. 2005. V. 5. No. 1. P. 269–282.
- [Suchorsky, Rand, 2009] *Suchorsky M., Rand R.* Three oscillator model of the heartbeat generator // Commun. Nonlinear Sci. Numer. Simulat. 2009. V. 14. P. 2434–2449.
- [Zeeman, 1972] *Zeeman E. C.* Differential equations for the heartbeat and nerve impulse // Towards a Theoretical Biology. 1972. V. 4. P. 8–67.

NONCLASSIC RELAXATION OSCILLATIONS AND BLUE SKY BIFURCATION IN CARDIORITHMS MODELING

S. D. Glyzin¹, A. Yu. Kolesov¹, N. H. Rozov²

¹ P. G. Demidov Yaroslavl State University (YSU)

² Lomonosov Moscow State University (MSU)

New class of systems of ordinary differential equations with one fast and two or more slow variables is considered. It is shown that under certain technical conditions on right-hand side functions of the model it can exhibit so-called nonclassic relaxation oscillations and blue-

sky bifurcation. The term nonclassic relaxation oscillations is used to denote oscillations with slow components asymptotically close to certain time-discontinuous functions, and fast component is of δ -like form. We present a map responsible for the attractor of the system (Poincare map along orbits on a special section) that makes it possible to prove existence of mentioned bifurcation. Cases when blue-sky catastrophe leads to relaxation cycle of stable two-dimensional torus are studied. Also considered is emergence of hyperbolic attractor (Smale–Williams attractor) in the system. An attempt of applying considered systems to cardiac rhythm simulations is described.

Keywords: singularly perturbed system, relaxation cycle, asymptotic behavior, stability, blue sky catastrophe, non-classical relaxation oscillations.

Glyzin Sergey Dmitrievich — doctor of physical and mathematical sciences, professor, glyzin@uni-yar.ac.ru

Kolesov Andrey Yurevich — doctor of physical and mathematical sciences, professor, kolesov@uni-yar.ac.ru

Rozov Nicolay Hristovich — doctor of physical and mathematical sciences, professor, fpo.mgu@mail.ru

РЕЖИМЫ ДВУХКЛАСТЕРНОЙ СИНХРОНИЗАЦИИ В ПОЛНОСВЯЗНОЙ СЕТИ РЕЛАКСАЦИОННЫХ ОСЦИЛЛЯТОРОВ

С. Д. Глызин¹, А. Ю. Колесов¹, Н. Х. Розов²

¹ Ярославский государственный университет им. П. Г. Демидова (ЯрГУ)

² Московский государственный университет имени М. В. Ломоносова (МГУ)

Статья посвящена проблеме математического моделирования нейронной активности. Предлагаются новые классы сингулярно возмущённых дифференциально-разностных уравнений с запаздыванием вольтерровского типа, с помощью которых описывается функционирование как отдельного нейрона, так и нейронных сетей. Проводится исследование аттракторов полносвязной системы импульсных нейронов при неограниченном увеличении числа её элементов. Для изучения периодических решений типа многокластерной синхронизации используются некоторые специальные приёмы, сводящие проблему существования и устойчивости циклов к анализу вспомогательной системы обыкновенных дифференциальных уравнений с импульсным воздействием. На этом пути устанавливается, что при увеличении числа элементов сети количество сосуществующих в ней устойчивых решений с двухкластерной синхронизацией растёт, т. е. имеет место известное явление буферности.

Исследование выполнено при поддержке Российского научного фонда (проект № 14-21-00158).

Ключевые слова: релаксационные осцилляторы, полносвязная сеть, кластерная динамика, асимптотическое поведение, устойчивость, феномен буферности.

1. ПОСТАНОВКА ЗАДАЧИ

Теоретической основой статьи служит общий подход к исследованию релаксационных колебаний в многомерных системах обыкновенных дифференциальных уравнений, берущий начало с работы Л. С. Понтрягина и Е. Ф. Мищенко (1955). К настоящему времени проведена целая серия исследований (см. [Глызин и др., 2011а, б, 2012а, б, 2013а, б; Колесов, Колесов, 1993; Колесов, Розов, 2010, 2011, 2012; Колесов и др., 1997, 2010]), связанных с переносом на уравнения с запаздыванием известных асимптотических методов [Мищенко, Розов, 1955; Мищенко и др., 1995]. Результаты этих работ существенно используются в данной статье. В частности, в ней анализируется некоторая система сингулярно возмущённых скалярных нелинейных дифференциально-разностных уравнений, моделирующая электрическую активность полносвязной сети импульсных нейронов с электрической связью. Рассматриваются вопросы о существовании и устойчивости в этой системе релаксационных циклов со специальными свойствами.

Приступим к описанию объекта дальнейшего анализа. Будем считать, что электрическая активность отдельного нейрона моделируется, построенным в соответствии с методикой книги [Кашенко, Майоров, 2009], дифференциально-разностным уравнением

$$\dot{u} = \lambda[f(u(t-h)) - g(u(t-1))]u. \quad (1)$$

Глызин Сергей Дмитриевич — заведующий кафедрой, доктор физико-математических наук, профессор, glyzin@uniyar.ac.ru

Колесов Андрей Юрьевич — доктор физико-математических наук, профессор, kolesov@uniyar.ac.ru

Розов Николай Христович — декан, доктор физико-математических наук, профессор, член-корреспондент РАН, fro.mgu@mail.ru

Здесь $u(t) > 0$ — мембранный потенциал нейрона, параметр $\lambda > 0$, характеризующий скорость протекания электрических процессов в системе, предполагается большим, а параметр h фиксирован и принадлежит интервалу $(0, 1)$. Относительно фигурирующих в выражении (1) функций $f(u)$, $g(u) \in C^2(\mathbb{R}_+)$, $\mathbb{R}_+ = \{u \in \mathbb{R} : u \geq 0\}$, предполагаем, что они обладают свойствами:

$$\begin{aligned} f(0) &= 1, \quad g(0) = 0, \quad f(u) = -a_0 + O(u^{-1}), \quad uf'(u) = O(u^{-1}), \\ u^2 f''(u) &= O(u^{-1}), \quad g(u) = b_0 + O(u^{-1}), \quad ug'(u) = O(u^{-1}), \\ u^2 g''(u) &= O(u^{-1}) \quad \text{при } u \rightarrow +\infty, \end{aligned} \quad (2)$$

где a_0, b_0 — положительные константы.

Уравнение (1) подробно исследовано в статье [Глызин и др., 2013а], а её упрощённый аналог

$$\dot{u} = \lambda f(u(t-1))u, \quad (3)$$

где функция $f(u)$ удовлетворяет перечисленным выше условиям (2), рассмотрен в работах [Колесов, Розов, 2010; Колесов и др., 2010]. Уравнение (3) получается из выражения (1) при $h=1$ и при переобозначениях

$$f(u) - g(u) \rightarrow f(u), \quad a_0 + b_0 \rightarrow a.$$

Для функции $f(u)$ будем считать выполненным следующее дополнительное требование:

$$a > 1, \quad (4)$$

которое имеет вполне понятную нейродинамическую интерпретацию, поскольку эквивалентно условию существования у уравнения (3) при достаточно большом λ устойчивого релаксационного цикла, соответствующего режиму генерации спайков (см. [Колесов, Розов, 2010; Колесов и др., 2010]).

В данной работе исследуются полностью связанные системы с электрической связью между нейронами, каждый из которых в отдельности моделируется уравнением (3)

$$\dot{u}_j = d \sum_{i=1}^m (u_i - u_j) + \lambda f(u_j(t-1))u_j, \quad j = 1, \dots, m. \quad (5)$$

Здесь $\lambda \gg 1$; $d = \text{const} > 0$; функция $f(u)$ удовлетворяет условиям (2), (4).

Математическое исследование модели (5) будем проводить при условии, что $\lambda \gg 1$. Система (5) допускает, очевидно, так называемый однородный или синхронный цикл

$$u_1 \equiv \dots \equiv u_m = u_*(t, \lambda), \quad (6)$$

где $u_*(t, \lambda)$ — устойчивое периодическое решение уравнения (3) (асимптотика этого решения построена в статье [Колесов и др., 2010]).

Не претендуя на полноту анализа, ввиду сложности системы (5) ограничимся изучением специальных её периодических решений — так называемых режимов двухкластерной синхронизации.

Для описания упомянутых режимов предположим, что совокупность индексов $1 \leq j \leq m$ разбита на два непересекающихся множества Σ_1, Σ_2 , т.е.

$$\{1, 2, \dots, m\} = \Sigma_1 \cup \Sigma_2. \quad (7)$$

Далее, обозначим через k и $m-k$ количества элементов в множествах Σ_1 и Σ_2 соответственно. Тогда, очевидно, система (5) допускает решения с компонентами

$$u_j = w_1(t) \text{ при } j \in \Sigma_1, \quad u_j = w_2(t) \text{ при } j \in \Sigma_2, \quad w_1(t) \neq w_2(t), \quad (8)$$

где переменные w_1, w_2 удовлетворяют системе

$$\begin{cases} \dot{w}_1 = d(k w_1 + (m-k)w_2 - m w_1) + \lambda f(w_1(t-1))w_1, \\ \dot{w}_2 = d(k w_1 + (m-k)w_2 - m w_2) + \lambda f(w_2(t-1))w_2. \end{cases} \quad (9)$$

Если же, в свою очередь, система (9) имеет периодическое решение, то отвечающее ему решение (8) исходной системы (5) назовём периодическим режимом двухкластерной синхронизации.

Все допустимые типы периодических решений двухкластерной синхронизации, отвечающие произвольным разбиениям (7), будем кодировать с помощью m -мерных символьных векторов вида

$$(\theta_1, \theta_2, \dots, \theta_m) : \theta_j = A \text{ или } B, \quad j = 1, \dots, m. \quad (10)$$

Предполагаем, что j -я координата вектора (10) равна A или B в случае $j \in \Sigma_1$ или $j \in \Sigma_2$ соответственно. При таком способе кодирования однородному циклу (6) будут соответствовать сразу два вектора (A, A, \dots, A) и (B, B, \dots, B) . В связи с этим данные кодировки считаем совпадающими.

Обращаем внимание, что всем векторам (10) с наперёд заданными количествами элементов A и B равными k и $m-k$, отвечает одна и та же система (9). Как будет показано ниже, при достаточно малых фиксированных значениях $d > 0$ и при всех $\lambda \gg 1$ эта система допускает два неоднородных цикла

$$\begin{aligned} C_1 : (w_1, w_2) &= (w_1^{(k, m-k)}(t), w_2^{(k, m-k)}(t)), \\ C_2 : (w_1, w_2) &= (w_2^{(m-k, k)}(t), w_1^{(m-k, k)}(t)), \end{aligned} \quad (11)$$

переходящих друг в друга при заменах $k \rightarrow m-k, w_1 \rightarrow w_2, w_2 \rightarrow w_1$. Так как порядок следования переменных w_1, w_2 для нас несущественен, то для определённости всегда будем считать, что функции $w_1(t), w_2(t)$ из соотношения (8) заданы, к примеру, равенствами

$$w_1(t) = w_1^{(k, m-k)}(t), \quad w_2(t) = w_2^{(k, m-k)}(t). \quad (12)$$

Отметим, что при описанных выше способах кодирования и отыскания компонент $w_1(t), w_2(t)$ имеет место взаимно однозначное соответствие между множеством всех периодических режимов двухкластерной синхронизации и множеством символьных векторов (10). Таким образом, с учётом однородного цикла (6) существует ровно $2^m - 1$ двухкластерных режимов. Как выяснится в дальнейшем, при достаточно малых значениях параметра d все они экспоненциально орбитально устойчивы.

Отметим ещё одно полезное наблюдение. Несложная проверка показывает, что система (5) инвариантна относительно замен

$$(u_1, u_2, \dots, u_m) \rightarrow (u_{j_1}, u_{j_2}, \dots, u_{j_m}), \quad (13)$$

где (j_1, j_2, \dots, j_m) — произвольная перестановка набора индексов $\{1, 2, \dots, m\}$. А это значит, что если мы найдём циклы (8), (12), соответствующие символьным векторам

$$\underbrace{(A, \dots, A)}_k \underbrace{(B, \dots, B)}_{m-k} \quad (14)$$

при каждом k , $1 \leq k \leq m-1$, и применим к ним преобразования (13), то в итоге будут исчерпаны все периодические режимы двухкластерной синхронизации, отличные от однородного цикла.

Основной результат состоит в том, что при подходящем уменьшении d и при всех $\lambda \gg 1$ эта система с учётом устойчивого цикла (6) имеет как минимум $2^m - 1$ экспоненциально орбитально устойчивых неоднородных периодических движений типа двухкластерной синхронизации. Учитывая, что размерность m системы (5) можно выбирать сколь угодно большой, то это означает реализуемость в рассматриваемой нейронной модели феномена буферности.

Для обоснования этого результата будем пользоваться доказанными в работах [Глызин и др., 2011а, б, 2012а; Колесов, Розов, 2010, 2012] для случая диффузионной связи основными утверждениями, модифицируя их для нашей задачи. Ниже приводятся базовые теоремы, на основе которых удаётся найти асимптотические формулы устойчивых периодических решений системы (5).

2. ОСНОВНЫЕ УТВЕРЖДЕНИЯ

Для удобства дальнейшего асимптотического анализа перейдем в (5) к новым переменным x, y_1, \dots, y_{m-1} по формулам

$$u_1 = \exp\left[\frac{x}{\varepsilon}\right], \quad u_j = \exp\left[\frac{x}{\varepsilon} + \sum_{k=1}^{j-1} y_k\right], \quad j=2, \dots, m, \quad \varepsilon = \frac{1}{\lambda}. \quad (15)$$

В результате получаем систему

$$\dot{x} = \varepsilon d \sum_{i=1}^m (\exp y_i - 1) + F(x(t-1), \varepsilon), \quad (16)$$

$$\begin{aligned} \dot{y}_j = d \left[\sum_{r=1}^j \exp\left[-\sum_{s=r}^j y_s\right] + \sum_{r=j+2}^m \exp\left[\sum_{s=j+1}^{r-1} y_s\right] - \sum_{r=1}^{j-1} \exp\left[-\sum_{s=r}^{j-1} y_s\right] - \right. \\ \left. - \sum_{r=j+1}^m \exp\left[\sum_{s=j}^{r-1} y_s\right] \right] + G_j(x(t-1), y_1(t-1), \dots, y_j(t-1), \varepsilon), \quad j=1, \dots, m-1, \end{aligned} \quad (17)$$

где функции F, G_j задаются соотношениями

$$\begin{aligned} F(x, \varepsilon) = f(\exp(x/\varepsilon)), \quad G_j(x, y_1, \dots, y_j, \varepsilon) = \\ = \frac{1}{\varepsilon} \left[f\left[\exp\left[\frac{x}{\varepsilon} + \sum_{k=1}^j y_k\right]\right] - f\left[\exp\left[\frac{x}{\varepsilon} + \sum_{k=1}^{j-1} y_k\right]\right] \right], \quad j=1, \dots, m-1. \end{aligned} \quad (18)$$

Отметим, что суммы в соотношении (17), у которых нижний индекс больше верхнего, считаются равными нулю.

Фиксируем постоянную σ_0 , подчинённую требованиям $0 < \sigma_0 < a-1$, и рассмотрим банахово пространство \mathcal{F} непрерывных при $-1-\sigma_0 \leq t \leq -\sigma_0$ вектор-функций $\varphi(t) = (\varphi_1(t), \dots, \varphi_m(t))$ с нормой

$$\|\varphi\|_{\mathcal{F}} = \max_{1 \leq j \leq m} \left(\max_{-1-\sigma_0 \leq t \leq -\sigma_0} |\varphi_j(t)| \right). \quad (19)$$

Всюду ниже нас будут интересовать решения системы (16), (17) с начальными условиями из множества

$$\mathcal{S} = \left\{ \varphi(t) = (\varphi_1(t), \dots, \varphi_m(t)) : \varphi_1 \in \mathcal{S}_1, \varphi_2 \in \mathcal{S}_2, \dots, \varphi_m \in \mathcal{S}_m \right\} \subset \mathcal{F}. \quad (20)$$

Здесь через \mathcal{S}_1 обозначено замкнутое, ограниченное и выпуклое множество функций $\varphi_1(t)$, удовлетворяющих требованиям $-q_1 \leq \varphi_1(t) \leq -q_2$, $\varphi_1(-\sigma_0) = -\sigma_0$, $q_1 > \sigma_0$, $q_2 \in (0, \sigma_0)$, а в качестве $\mathcal{S}_2, \dots, \mathcal{S}_m$ взяты произвольные замкнутые и ограниченные подмножества пространства $C[-1-\sigma_0, -\sigma_0]$.

Формулировка строгих результатов об автоволновых режимах системы (16), (17) требует некоторых подготовительных построений. В связи с этим введём в рассмотрение решение

$$(x_\varphi(t, \varepsilon), y_{1,\varphi}(t, \varepsilon), \dots, y_{m-1,\varphi}(t, \varepsilon)), \quad t \geq -\sigma_0 \quad (21)$$

упомянутой системы, отвечающее произвольному начальному условию $\varphi(t)$ из множества \mathcal{S} . Рассмотрим также второй положительный корень $t = T_\varphi$ уравнения $x_\varphi(t - \sigma_0, \varepsilon) = -\sigma_0$ (в случае, когда он существует) и на множестве (20) определим оператор $\Pi_\varepsilon : \mathcal{S} \rightarrow \mathcal{F}$ посредством равенства

$$\begin{aligned} \Pi_\varepsilon(\varphi) = & (x_\varphi(t + T_\varphi, \varepsilon), y_{1,\varphi}(t + T_\varphi, \varepsilon), \dots, y_{m-1,\varphi}(t + T_\varphi, \varepsilon)), \\ & -1 - \sigma_0 \leq t \leq -\sigma_0. \end{aligned} \quad (22)$$

Помимо (22) нам потребуется ещё оператор $\Pi_0 : \mathcal{S} \rightarrow \mathcal{F}$, который зададим формулой

$$\begin{aligned} \Pi_0(\varphi) = & (x_0(t), y_1^0(t + T_0, z), \dots, y_{m-1}^0(t + T_0, z))|_{z=(\varphi_2(-\sigma_0), \dots, \varphi_m(-\sigma_0))}, \\ & -1 - \sigma_0 \leq t \leq -\sigma_0. \end{aligned} \quad (23)$$

Здесь, как и в статье [Колесов, Розов, 2010; Колесов и др., 2010], $T_0 = 2 + a + a^{-1}$; $x_0(t)$ — периодическая функция

$$x_0(t) = \begin{cases} t & \text{при } 0 \leq t \leq 1, \\ 1 - a(t-1) & \text{при } 1 \leq t \leq t_0 + 1, \quad x_0(t + T_0) \equiv x_0(t), \\ -a + t - t_0 - 1 & \text{при } t_0 + 1 \leq t \leq T_0. \end{cases} \quad (24)$$

Что же касается компонент $y_1^0(t, z), \dots, y_{m-1}^0(t, z)$, зависящих от вектора $z = (z_1, \dots, z_{m-1}) \in \mathbb{R}^{m-1}$, то при $-1 - \sigma_0 \leq t \leq T_0 - \sigma_0$ они удовлетворяют импульсной задаче Коши

$$\begin{aligned} \dot{y}_j = & d \left[\sum_{r=1}^j \exp \left(-\sum_{s=r}^j y_s \right) + \sum_{r=j+2}^m \exp \left(\sum_{s=j+1}^{r-1} y_s \right) - \right. \\ & \left. - \sum_{r=1}^{j-1} \exp \left(-\sum_{s=r}^{j-1} y_s \right) - \sum_{r=j+1}^m \exp \left(\sum_{s=j}^{r-1} y_s \right) \right], \end{aligned} \quad (25)$$

$$\begin{aligned} y_j(1+0) &= y_j(1-0) - (1+a)y_j(0), \\ y_j(t_0+1+0) &= y_j(t_0+1-0) - (1+a^{-1})y_j(t_0), \quad j = 1, \dots, m-1, \end{aligned} \quad (26)$$

$$(y_1, \dots, y_{m-1})|_{t=-\sigma_0} = z, \quad z = (z_1, \dots, z_{m-1}), \quad (27)$$

где $t_0 = 1 + a^{-1}$ и, как и выше, суммы в соотношении (25), у которых нижний индекс больше верхнего, считаются равными нулю.

Отдельно остановимся на вопросе о корректности определения оператора (23). Главная проблема здесь заключается в том, что на промежутках времени $-\sigma_0 \leq t < 1$, $1 \leq t < t_0 + 1$ и $t_0 + 1 \leq t \leq T_0 - \sigma_0$ решение задачи Коши (25), (26), (27) удовлетворяет нелинейной системе (25), поэтому возникает вопрос о продолжимости решений последней в указанные промежутки, длины которых отнюдь не малы. Ответ на него даётся в следующем утверждении.

Лемма 1. *Решение $(y_1(t), \dots, y_{m-1}(t))$ системы (25) с произвольным начальным условием $(y_1, \dots, y_{m-1})|_{t=0} = (\bar{y}_1, \dots, \bar{y}_{m-1}) \in \mathbb{R}^{m-1}$ определено на полуоси $t \geq 0$ и стремится к нулю при $t \rightarrow +\infty$.*

Для проверки требуемого факта достаточно заметить, что любое решение системы (25) записывается в виде

$$y_j(t) = \ln \left| \frac{\xi_{j+1}(t)}{\xi_j(t)} \right| \quad j = 1, \dots, m-1, \quad (28)$$

где $(\xi_1(t), \dots, \xi_m(t))$ — произвольное решение линейной системы

$$\dot{\xi}_j = d \left[\sum_{i=1}^m \xi_i - m \xi_j \right], \quad j = 1, \dots, m \quad (29)$$

из инвариантного конуса K векторов с положительными координатами. Из свойств системы (29) и замены (28) и вытекает утверждение леммы 1.

Другая проблема, связанная с корректностью оператора (23), в том, что функции $y_j^0(t, z)$, $j = 1, \dots, m-1$ являются разрывными в точках $t = 1$ и $t = t_0 + 1$, где согласно (26) они претерпевают конечные скачки. Однако в силу неравенства (4) и оценки $\sigma_0 < a - 1$ функции $y_j^0(t + T_0, z)$, $j = 1, \dots, m-1$ оказываются непрерывными на нужном отрезке $-1 - \sigma_0 \leq t \leq -\sigma_0$, поскольку в этом случае $T_0 - 1 - \sigma_0 > t_0 + 1$. Тем самым, условие (4) гарантирует выполнение требуемого включения $\Pi_0(\varphi) \in F$ при $\forall \varphi \in S$.

Завершая описание подготовительной части, рассмотрим производные Фреше $\partial_\varphi \Pi_\varepsilon(\varphi)$, $\partial_\varphi \Pi_0(\varphi)$ операторов (22), (23) по переменной φ . В данном случае эти производные представляют собой линейные операторы, действующие в пространстве

$$\mathcal{F}_0 = \left\{ \mathbf{g}_0(t) = (\mathbf{g}_{1,0}(t), \mathbf{g}_{2,0}(t), \dots, \mathbf{g}_{m,0}(t)) \in \mathcal{F} : \mathbf{g}_{1,0}(-\sigma_0) = 0 \right\}$$

с нормой (19), а результаты их применения к произвольному элементу $\mathbf{g}_0(t)$ из \mathcal{F}_0 задаются соответственно равенствами

$$\begin{aligned} \partial_\varphi \Pi_\varepsilon(\varphi) \mathbf{g}_0 &= (\mathbf{g}_1(t + T_\varphi, \varepsilon), \dots, \mathbf{g}_m(t + T_\varphi, \varepsilon)) - \\ &- l(\mathbf{g}_0) \left(\dot{\mathbf{x}}_\varphi(t + T_\varphi, \varepsilon), \dot{y}_{1,\varphi}(t + T_\varphi, \varepsilon), \dots, \dot{y}_{m-1,\varphi}(t + T_\varphi, \varepsilon) \right), \\ &- 1 - \sigma_0 \leq t \leq -\sigma_0; \end{aligned} \quad (30)$$

$$\begin{aligned} & \partial_{\varphi} \Pi_0(\varphi)g_0 = \\ & = \left[\begin{array}{c} 0, \sum_{s=1}^{m-1} \frac{\partial y_1^0}{\partial z_s}(t+T_0, z)g_{s+1,0}(-\sigma_0), \sum_{s=1}^{m-1} \frac{\partial y_2^0}{\partial z_s}(t+T_0, z)g_{s+1,0}(-\sigma_0), \dots \\ \dots, \sum_{s=1}^{m-1} \frac{\partial y_{m-1}^0}{\partial z_s}(t+T_0, z)g_{s+1,0}(-\sigma_0), \end{array} \right], \\ & z = (\varphi_2(-\sigma_0), \dots, \varphi_m(-\sigma_0)), \quad -1-\sigma_0 \leq t \leq -\sigma_0. \end{aligned} \quad (31)$$

Здесь $g(t, \varepsilon) = (g_1(t, \varepsilon), \dots, g_m(t, \varepsilon))$, $g(t, \varepsilon) = g_0(t)$ при $-1-\sigma_0 \leq t \leq -\sigma_0$ — решение линейной системы, получающейся из выражений (16), (17) при линеаризации на решении (21), а функционал $l: \mathcal{F}_0 \rightarrow \mathbb{R}$ определён соотношением

$$l(g_0) = g_1 \frac{T_{\varphi} - \sigma_0, \varepsilon}{\dot{x}_{\varphi}(T_{\varphi} - \sigma_0, \varepsilon)}. \quad (32)$$

Естественно возникающий вопрос о связи между операторами (22) и (23) решается следующей теоремой.

Теорема 1 (о C^1 -сходимости). Пусть выполнено условие (4) и множество S выбрано описанным выше способом. Тогда найдётся такое достаточно малое $\varepsilon_0 = \varepsilon_0(S) > 0$, что при всех $0 < \varepsilon \leq \varepsilon_0$ оператор Π_{ε} определён на S и удовлетворяет предельным равенствам

$$\left. \begin{aligned} \limsup_{\varepsilon \rightarrow 0} \sup_{\varphi \in S} \|\Pi_{\varepsilon}(\varphi) - \Pi_0(\varphi)\|_{\mathcal{F}} &= 0, \\ \limsup_{\varepsilon \rightarrow 0} \sup_{\varphi \in S} \|\partial_{\varphi} \Pi_{\varepsilon}(\varphi) - \partial_{\varphi} \Pi_0(\varphi)\|_{\mathcal{F}_0 \rightarrow \mathcal{F}_0} &= 0. \end{aligned} \right\} \quad (33)$$

Доказательство этой теоремы опустим, поскольку в случае $m = 2$ оно приведено в статье [Колесов, Розов, 2010]. Переход же от двумерного случая к значениям $m > 2$ носит чисто технический характер.

Остановимся на одном важном следствии из C^1 -сходимости, касающемся существования и устойчивости периодических решений системы (16), (17). В связи с этим обратим внимание, что в силу (25), (26), (27) предельный оператор (23) является надстройкой над соответствующим $(m - 1)$ -мерным отображением

$$z \rightarrow \Phi(z) = \left. \begin{array}{c} \text{def} \\ y_1^0(t, z), y_2^0(t, z), \dots, y_{m-1}^0(t, z) \end{array} \right|_{t=T_0-\sigma_0}, \quad (34)$$

где $z = (\varphi_2(-\sigma_0), \dots, \varphi_m(-\sigma_0))$.

Справедлив следующий результат, позволяющий сопоставить любой неподвижной точке $z = z_*$ этого отображения релаксационный цикл системы (16), (17).

Теорема 2 (о соответствии). Каждой неподвижной точке $z = z_*$ отображения (34), удовлетворяющей условию $\det(I - \Phi'(z_*)) \neq 0$, где I — единичная матрица, соответствует релаксационный цикл системы (16), (17). Этот цикл существует при всех достаточно малых $\varepsilon > 0$ и является экспоненциально орбитально устойчивым (неустойчивым) при $r_* < 1 (> 1)$, где r_* — спектральный радиус матрицы $\Phi'(z_*)$.

Доказательство данного утверждения также опустим, поскольку оно может быть легко получено модификацией соответствующего утверждения из работ [Глызин и др., 2012а; Колесов, Розов, 2010] для системы связанных осцилляторов с диффузионным взаимодействием.

Теорема 2 сводит интересующую нас проблему кластерной синхронизации системы (16), (17) к поиску неподвижных точек отображения (34). Прежде, чем перейти к вопросу о количестве и устойчивости таких точек заметим, что это отображение может быть приведено к инвариантному виду, не зависящему от начального момента времени $t = -\sigma_0$.

$$z \rightarrow \Phi_0(z) = \left(y_1^0(t, z), y_2^0(t, z), \dots, y_{m-1}^0(t, z) \right) \Big|_{t=T_0}, \quad (35)$$

где $(y_1^0(t, z), y_2^0(t, z), \dots, y_{m-1}^0(t, z))$ — решение аналогичной (25), (26), (27) задачи Коши для системы (25), (26) с начальным условием

$$(y_1, \dots, y_{m-1}) \Big|_{t=0} = z, \quad z = (z_1, \dots, z_{m-1}). \quad (36)$$

Подчеркнём, что в силу леммы 1 это отображение заведомо определено во всем пространстве \mathbb{R}^{m-1} .

3. ЦИКЛЫ ДВУХКЛАСТЕРНОЙ СИНХРОНИЗАЦИИ

Обратимся теперь к интересующему нас вопросу о существовании и устойчивости у системы (5) циклов двухкластерной синхронизации. Как уже было сказано выше, данная проблема сводится к исследованию лишь специальных циклов этого семейства, закодированных векторами (14). Для таких циклов очевидным образом имеем

$$y_j = \ln \left| \frac{\dot{u}_j}{u_j} \right| \equiv 0 \quad \text{при } j = 1, \dots, m-1, \quad j \neq k; \quad y_k \neq 0. \quad (37)$$

Подставляя, далее, соотношения (37) в выражения (25)–(27), приходим к выводу, что от исходного отображения (35) «отщепляется» одномерное отображение вида

$$z_k \rightarrow \Phi_{0,k}(z_k) = y_k^0(t, z_k) \Big|_{t=T_0}, \quad (38)$$

где $y_k^0(t, z_k)$ — решение импульсной задачи Коши

$$\begin{aligned} \dot{y} &= d(k \cdot \exp(-y) + m - 2k - (m - k) \cdot \exp y), \\ y(1+0) &= y(1-0) - (1+a)y(0), \\ y(t_0+1+0) &= y(t_0+1-0) - (1+a^{-1})y(t_0), \end{aligned} \quad (39)$$

$$y \Big|_{t=0} = z_k. \quad (40)$$

Добавим, что получившееся одномерное отображение в силу своего происхождения отвечает за существование и устойчивость циклов вспомогательной системы (9).

Как и во всех предыдущих случаях, отыскание неподвижных точек отображения (38) проведём в предположении о малости d . Следуя изложенной в работах [Глызин и др., 2011б; Колесов, Розов, 2010] методике, подставим в выражение (40) соотношение

$$z_k = \frac{1}{a} \ln \frac{1}{d} + v, \quad v = \text{const} \in \mathbb{R} \quad (41)$$

и обозначим через $y_k(t, v, d)$ решение получившейся задачи Коши (39)–(41). Проводя соответствующие подсчёты, убеждаемся, что при $d \rightarrow 0$ справедливы следующие асимптотические представления:

$$y_k = \frac{1}{a} \ln \frac{1}{d} + v + O(d^{1-a^{-1}}) \text{ при } 0 \leq t < 1; \quad (42)$$

$$y_k = -\ln \frac{1}{d} + \ln[k(t-1) + \exp(-av)] + O(d^{1-a^{-1}}) \text{ при } 1 \leq t < t_0 + 1; \quad (43)$$

$$y_k = \frac{1}{a} \ln \frac{1}{d} + \psi_k(v) + O(d^{1-a^{-1}}) \text{ при } t_0 + 1 \leq t \leq T_0, \quad (44)$$

где

$$\psi_k(v) = \ln[kt_0 + \exp(-av)] - (1 + a^{-1}) \ln[k(t_0 - 1) + \exp(-av)]. \quad (45)$$

Соотношения (42)–(44) свидетельствуют о том, что после выполнения в отображении (38) замены (41) оно допускает при $d \rightarrow 0$ (в C^1 -метрике на любом конечном отрезке изменения v) предел

$$v \rightarrow \psi_k(v). \quad (46)$$

Анализ отображения (46) показывает, что при выполнении неравенства $kt_0 > (1 + a^{-1})k(t_0 - 1)$, которое эквивалентно (4), отображение (46) имеет единственную экспоненциально устойчивую неподвижную точку $v = v_{*,k}$. В исходном отображении (38) этой неподвижной точке отвечает устойчивая неподвижная точка

$$z_k = z_k^*, \quad z_k^* = \frac{1}{a} \ln \frac{1}{d} + v_{*,k} + O(d^{1-a^{-1}}), \quad d \rightarrow 0, \quad (47)$$

а в системе (9) — устойчивый релаксационный цикл C_1 (см. (11)).

Следует также заметить, что в силу инвариантности системы (39) по отношению к преобразованиям $k \rightarrow m - k$, $y \rightarrow -y$ наряду с точкой (47) отображение (38) допускает также устойчивую неподвижную точку $z_k = z_k^{**}$ с асимптотикой

$$z_k^{**} = -\frac{1}{a} \ln \frac{1}{d} - v_{*,m-k} + O(d^{1-a^{-1}}), \quad d \rightarrow 0.$$

Этой неподвижной точке в системе (9) соответствует цикл C_2 (см. (11)), не представляющий для нас интереса, поскольку выше мы договорились определять компоненты $w_1(t)$, $w_2(t)$ из соотношения (8) по правилам (12).

Итак, для каждого натурального k , $1 \leq k \leq m - 1$, удалось найти цикл типа двухкластерной синхронизации, ассоциированный с символьным вектором (14). Согласно развитой нами теории за устойчивость этого цикла отвечает расположение спектра матрицы Якоби $\Phi'_0(\bar{z}_{(k)})$ отображения (35) в неподвижной точке $\bar{z}_{(k)} = (z_1, \dots, z_{m-1})$: $z_j = 0$ при $j \neq k$, $z_k = z_k^*$, где z_k^* — неподвижная точка (47) отображения (38).

Одно собственное значение матрицы $\Phi'_0(\bar{z}_{(k)})$ заранее известно — это мультипликатор неподвижной точки (47), задающийся равенством

$$\mu = \left. \frac{d\Phi_{0,k}}{dz_k} \right|_{z_k=z_k^*} = \left. \frac{d\psi_k}{dv} \right|_{v=v_{*,k}} + O(d^{1-a^{-1}}), \quad d \rightarrow 0$$

и в силу отмеченных выше свойств отображения (46) принадлежащий интервалу $(0, 1)$. Другие же собственные значения данной матрицы совпадают со спектром оператора сдвига Γ за время от $t=0$ до $t=T_0$ по решениям импульсной системы:

$$\dot{g}_j = d \begin{pmatrix} -\sum_{r=1}^j \sum_{s=r}^j g_s + \sum_{r=j+2s=j+1}^m \sum_{r-1}^{r-1} g_s + \exp(y_k^*) \sum_{r=k+1}^m \sum_{s=j+1}^{r-1} g_s + \\ + \sum_{r=1}^{j-1} \sum_{s=r}^{j-1} g_s - \sum_{r=j+1s=j}^k \sum_{r-1}^{r-1} g_s - \exp(y_k^*) \sum_{r=k+1}^m \sum_{s=j}^{r-1} g_s \end{pmatrix}, \quad 1 \leq j < k; \quad (48)$$

$$\dot{g}_j = d \begin{pmatrix} -\exp(-y_k^*) \sum_{r=1}^k \sum_{s=r}^j g_s - \sum_{r=k+1s=r}^j \sum_{r-1}^{r-1} g_s + \sum_{r=j+2s=j+1}^m \sum_{r-1}^{r-1} g_s + \\ + \exp(-y_k^*) \sum_{r=1}^k \sum_{s=k}^{j-1} g_s + \sum_{r=k+1s=r}^{j-1} \sum_{r-1}^{j-1} g_s - \sum_{r=j+1s=j}^m \sum_{r-1}^{r-1} g_s \end{pmatrix}, \quad k < j \leq m-1; \quad (49)$$

$$\begin{aligned} g_j(1+0) &= g_j(1-0) - (1+a)g_j(0), \quad g_j(t_0+1+0) = \\ &= g_j(t_0+1-0) - (1+a^{-1})g_j(t_0), \quad j=1, \dots, m-1, \quad j \neq k. \end{aligned} \quad (50)$$

Здесь функция y_k^* является решением задачи Коши (39), (40) при $z_k = z_k^*$ и вследствие этого допускает асимптотические представления вида (42)–(44) (в которых v и $\Psi_k(v)$ необходимо заменить на неподвижную точку $v_{*,k}$ отображения (46)).

Для вычисления спектра введённого выше оператора Γ дополним систему (48)–(50) произвольными независимыми от d начальными условиями $g_j|_{t=0} = g_j^0$, $j=1, \dots, m-1$, $j \neq k$ и предпримем асимптотическое интегрирование получившейся задачи Коши.

Как обычно, начнём с промежутка $0 \leq t < 1$. Опираясь на формулу (42), приходим к выводу, что в данном случае система (48), (49) записывается в виде

$$\dot{g}_j = d^{1-a^{-1}} \exp(v_{*,k}) \left[\sum_{r=k+1s=j+1}^m \sum_{s=k}^{r-1} g_s - \sum_{r=k+1s=j}^m \sum_{s=k}^{r-1} g_s \right] + o(d^{1-a^{-1}}) = \quad (51)$$

$$= -d^{1-a^{-1}}(m-k)g_j \exp(v_{*,k}) + o(d^{1-a^{-1}}), \quad j < k;$$

$$\dot{g}_j = O(d), \quad j > k, \quad d \rightarrow 0. \quad (52)$$

А отсюда, в свою очередь, выводим:

$$g_j = g_j^0 \left[1 - d^{1-a^{-1}} t(m-k) \cdot \exp(v_{*,k}) \right] + o(d^{1-a^{-1}}), \quad j < k;$$

$$g_j = g_j^0 + O(d), \quad j > k, \quad d \rightarrow 0. \quad (53)$$

Следует особо подчеркнуть, что для краткости символами $o(d^{1-a^{-1}})$, $O(d)$ в формулах (51)–(52) и в последующих формулах для \dot{g}_j , g_j обозначаются некоторые линейные формы переменных g_s , $s=1, \dots, m-1$, $s \neq k$ (в случае \dot{g}_j) или

g_s^0 , $s=1, \dots, m-1$, $s \neq k$ (в случае g_j) с коэффициентами, зависящими от t, d . Что же касается упомянутых коэффициентов, то именно они равномерно по t имеют указанные порядки малости при $d \rightarrow 0$.

При $1 \leq t < t_0 + 1$, учитывая равенство (43) и уже установленные формулы (53), получаем задачу Коши

$$\dot{g}_j = O(d), \quad j < k;$$

$$\begin{aligned} \dot{g}_j &= \frac{1}{k(t-1) + \exp(-av_{*,k})} \left[\sum_{r=1}^k \sum_{\substack{s=r \\ s \neq k}}^{j-1} g_s - \sum_{r=1}^k \sum_{\substack{s=r \\ s \neq k}}^j g_s \right] + O(d^{1-a^{-1}}) = \\ &= -\frac{kg_j}{k(t-1) + \exp(-av_{*,k})} + O(d^{1-a^{-1}}), \quad j > k; \end{aligned}$$

$$g_j \Big|_{t=1} = -g_j^0 \left[a + d^{1-a^{-1}} (m-k) \cdot \exp(v_{*,k}) \right] + o(d^{1-a^{-1}}), \quad j < k;$$

$$g_j \Big|_{t=1} = -ag_j^0 + O(d), \quad j > k, \quad d \rightarrow 0.$$

Несложный её анализ приводит к асимптотическим представлениям

$$g_j = -g_j^0 \left[a + d^{1-a^{-1}} (m-k) \cdot \exp(v_{*,k}) \right] + o(d^{1-a^{-1}}), \quad j < k;$$

$$g_j = -\frac{a \exp(-av_{*,k})}{k(t-1) + \exp(-av_{*,k})} g_j^0 + O(d^{1-a^{-1}}), \quad j > k, \quad d \rightarrow 0. \quad (54)$$

На заключительном участке $t_0 + 1 \leq t \leq T_0$, объединяя формулы (44) и (54), снова приходим к системе вида (51), (52), но уже с начальными условиями

$$g_j \Big|_{t=t_0+1} = g_j^0 \left[1 + a^{-1} d^{1-a^{-1}} (m-k) \cdot \exp(v_{*,k}) \right] + o(d^{1-a^{-1}}), \quad j < k;$$

$$g_j \Big|_{t=t_0+1} = g_j^0 \frac{d\psi_k}{dv} \Big|_{v=v_{*,k}} + O(d^{1-a^{-1}}), \quad j > k, \quad d \rightarrow 0, \quad (55)$$

где, напомним, $\psi_k(v)$ — функция (45). Проводя асимптотическое интегрирование задачи Коши (51), (52), (55), убеждаемся, что

$$g_j = g_j^0 \left[1 - d^{1-a^{-1}} (m-k)(t-t_0-1-a^{-1}) \cdot \exp(v_{*,k}) \right] + o(d^{1-a^{-1}}), \quad j < k;$$

$$g_j = g_j^0 \frac{d\psi_k}{dv} \Big|_{v=v_{*,k}} + O(d^{1-a^{-1}}), \quad j > k, \quad d \rightarrow 0. \quad (56)$$

Подводя итог, заметим, что интересующий нас оператор Γ задаётся соотношением

$$\left(g_1^0, \dots, g_{k-1}^0, g_{k+1}^0, \dots, g_{m-1}^0 \right) \rightarrow \left(g_1, \dots, g_{k-1}, g_{k+1}, \dots, g_{m-1} \right) \Big|_{t=T_0}. \quad (57)$$

Подставляя, далее, в соотношение (57) формулы (56), приходим к выводу, что его собственные значения допускают при $d \rightarrow 0$ асимптотику

$$\mu_j = 1 - (a - a^{-1}) d^{1-a^{-1}} (m-k) \cdot \exp(v_{*,k}) + o(d^{1-a^{-1}}), \quad j=1, \dots, k-1;$$

$$\mu_j = \left. \frac{d\psi_k}{dv} \right|_{v=v_{*,k}} + O(d^{1-a^{-1}}), \quad j = k, \dots, m-2$$

и вследствие этого удовлетворяют неравенствам $|\mu_j| < 1, \quad j = 1, \dots, m-2$.

Итак, все найденные периодические режимы системы (5) типа двухкластерной синхронизации, отличные от однородного цикла (6), при малых фиксированных d и при $\lambda \gg 1$ экспоненциально орбитально устойчивы. Что же касается однородного цикла, то он не является исключением, поскольку отвечающая ему неподвижная точка $z=0$ отображения (35) по-прежнему экспоненциально устойчива при всех $d > 0$.

Рассуждая аналогичным приведённому выше образом, приходим к выводу, что собственные значения матрицы Якоби $\Phi'_0(0)$ задаются равенствами $\mu_k = \mu(s)|_{s=s_k}, \quad k = 1, \dots, m-1$, где $\mu(s) = ((1+a)\exp s - 1)((1+a^{-1})\exp s - 1)\exp(-sT_0)$, а $s = s_k$ — ненулевые собственные значения матрицы системы (29). Функция $\mu(s)$ удовлетворяет следующим свойствам: $\mu(0) = 1, \quad \mu'(s) < 0$, при $\forall s \in (0, +\infty)$, $\mu(s) \rightarrow +\infty$ при $s \rightarrow +\infty$. Учитывая, что $s_k = -m, \quad k = 1, \dots, m-1$, имеем $\mu_k \in (0, 1), \quad k = 1, \dots, m-1$.

ЗАКЛЮЧЕНИЕ

Подводя итоги, отметим, что теорема 2 о соответствии и проделанный в пункте 3 асимптотический анализ отображения (35) приводят к следующему утверждению, являющемуся основным результатом данной статьи.

Теорема 3. Пусть параметр a удовлетворяет неравенству (4). Тогда для любых достаточно малых $d_2 > d_1 > 0$ найдётся такое достаточно большое $\lambda_0 = \lambda_0(d_1, d_2) > 0$, что при всех $d_1 \leq d \leq d_2, \quad \lambda \geq \lambda_0$ система (5) имеет как минимум 2^m экспоненциально орбитально устойчивых периодических движений.

Из этой теоремы следует, что при согласованном уменьшении d и увеличении параметров λ, m в системе наблюдается феномен буферности, а точнее, происходит неограниченное накапливание сосуществующих устойчивых циклов.

Отметим, что буферность наблюдается также и для ряда классов цепочек осцилляторов с однонаправленно связью (см. [Глызин и др., 2013б; 2014а–г]).

ЛИТЕРАТУРА

- [Глызин и др., 2011а] Глызин С. Д., Колесов А. Ю., Розов Н. Х. Релаксационные автоколебания в нейронных системах. I // Дифференц. уравнения. 2011. Т. 47. № 7. С. 919–932.
- [Глызин и др., 2011б] Глызин С. Д., Колесов А. Ю., Розов Н. Х. Релаксационные автоколебания в нейронных системах. II // Дифференц. уравнения. 2011. Т. 47. № 12. С. 1675–1692.
- [Глызин и др., 2012а] Глызин С. Д., Колесов А. Ю., Розов Н. Х. Релаксационные автоколебания в нейронных системах. III // Дифференц. уравнения. 2012. Т. 48. №. 2. С. 155–170.
- [Глызин и др., 2012б] Глызин С. Д., Колесов А. Ю., Розов Н. Х. Дискретные автоволны в нейронных системах // Журн. вычислит. математики и мат. физики. 2012. Т. 52. № 5. С. 840–858.

- [Глызин и др., 2013а] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Моделирование эффекта взрыва в нейронных системах // *Мат. заметки.* 2013. Т. 93. №. 5. С. 684–701.
- [Глызин и др., 2013б] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Периодические решения типа бегущих волн в кольцевых цепочках однонаправленно связанных уравнений // *Теоретическая и математическая физика.* 2013. Т. 175. №. 1. С. 62–83.
- [Глызин и др., 2014а] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Автоволновые процессы в континуальных цепочках однонаправленно связанных генераторов // *Избранные вопросы математической физики и анализа: сб. ст. К 90-летию со дня рождения академика Василия Сергеевича Владимировича* // *Тр. МИАН.* 2014. Т. 285. С. 89–106.
- [Глызин и др., 2014б] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Явление буферности в континуальных цепочках однонаправленно связанных генераторов // *Теор. и мат. физика.* 2014. Т. 181. № 2. С. 254–275.
- [Глызин и др., 2014в] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Явление буферности и хаос в кольцевых цепочках однонаправленно связанных генераторов // *Докл. Академии наук.* 2014. Т. 457. № 3. С. 278–281.
- [Глызин и др., 2014г] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Явление буферности в кольцевых цепочках однонаправленно связанных генераторов // *Изв. РАН. Сер. матем.* 2014. Т. 78. № 4. С. 73–108.
- [Кашенко, Майоров, 2009] *Кашенко С. А., Майоров В. В.* Модели волновой памяти. М.: Кн. дом «ЛИБРОКОМ», 2009. 288 с.
- [Колесов, Колесов, 1993] *Колесов А. Ю., Колесов Ю. С.* Релаксационные колебания в математических моделях экологии // *Тр. МИРАН.* 1993. Т. 199.
- [Колесов, Розов, 2010] *Колесов А. Ю., Розов Н. Х.* Дискретные автоволны в системах с запаздыванием из экологии // *Докл. Академии наук.* 2010. Т. 434. № 6. С. 735–738.
- [Колесов, Розов, 2011] *Колесов А. Ю., Розов Н. Х.* Теория релаксационных колебаний для уравнения Хатчинсона // *Матем. сб.* 2011. Т. 202. № 6. С. 51–82.
- [Колесов, Розов, 2012] *Колесов А. Ю., Розов Н. Х.* Автоволновые процессы в цепочках диффузионно связанных уравнений с запаздыванием // *УМН.* 2012. Т. 67. № 2(404). С. 109–156.
- [Колесов и др., 1997] *Колесов А. Ю., Мищенко Е. Ф., Розов Н. Х.* Реле с запаздыванием и его C^1 -аппроксимация // *Тр. Мат. ин-та РАН.* 1997. Т. 216. С. 126–153.
- [Колесов и др., 2010] *Колесов А. Ю., Мищенко Е. Ф., Розов Н. Х.* Об одной модификации уравнения Хатчинсона // *Журн. вычислит. математики и мат. физики.* 2010. Т. 50. № 12. С. 2099–2112.
- [Мищенко, Розов, 1955] *Мищенко Е. Ф., Розов Н. Х.* Дифференциальные уравнения с малым параметром и релаксационные колебания. М: Наука, 1975. 248 с.
- [Мищенко, Понтрягин, 1955] *Мищенко Е. Ф., Понтрягин Л. С.* Периодические решения систем дифференциальных уравнений, близкие к разрывным // *Докл. АН СССР.* 1955. Т. 102. № 5. С. 889–891.
- [Мищенко и др., 1995] *Мищенко Е. Ф., Колесов Ю. С., Колесов А. Ю., Розов Н. Х.* Периодические движения и бифуркационные процессы в сингулярно возмущённых системах. М.: Физматлит, 1995.

TWO-CLUSTER SYNCHRONIZATION REGIMES OF FULLY CONNECTED NETWORK OF RELAXATION OSCILLATORS

*S. D. Glyzin*¹, *A. Yu. Kolesov*¹, *N. H. Rozov*²

¹ P.G. Demidov Yaroslavl State University (P.G. Demidov YSU)

² M.V. Lomonosov Moscow State University (MSU)

The article is devoted to the mathematical modeling of neural activity. We propose new classes of singularly perturbed differential-difference equations of Volterra type with delay. On the base of these systems, models of a single neuron as well as neural networks are described. The neurons are coupled in an all-to-all manner. We study attractors of these systems in the case where the number of oscillators in the system increases indefinitely. In order to study periodic solutions of cluster synchronization type of this system, some special tricks are used. These methods reduce the existence and stability problems for cycles to the investigation of auxiliary system with impulse actions. Using this approach, we establish that the number of stable two-cluster solutions increases unboundedly as the number of elements of the network increases, that is, the well-known buffer phenomenon occurs.

Keywords: relaxation oscillator, fully connected network, clustering dynamics, asymptotic behaviour, stability, buffer phenomenon.

Glyzin Sergey Dmitrievich — head of chair, doctor of physical and mathematical sciences, professor, glyzin@uniyar.ac.ru

Kolesov Andrey Yurevich — doctor of physical and mathematical sciences, professor, kolesov@uniyar.ac.ru

Rozov Nicolay Hristovich — dean, doctor of physical and mathematical sciences, professor, fpo.mgu@mail.ru

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ ЭЛЕКТРОННЫХ СОСТОЯНИЙ В ДИОКСИДЕ ЦИРКОНИЯ С КИСЛОРОДНЫМИ ВАКАНСИЯМИ И ПРИМЕСЯМИ ИТТРИЯ

О. Н. Горшков, Н. В. Дербенева, К. Р. Мухаматчин, А. М. Сатанин

Нижегородский государственный университет им. Н. И. Лобачевского
(ННГУ им. Н. И. Лобачевского)

В работе приводятся результаты исследования электронных состояний в тонких плёнках на основе диоксида циркония (ZrO_2), стабилизированного иттрием (Y). Методом Кона-Шэма выполнены расчёты электронных состояний в допированном и недопированном диоксиде циркония. Изучена перестройка вакансионных состояний в зависимости от концентрации иттрия (степени допирования). При определённых условиях равновесия в подсистеме кислород-иттрий стабилизируется конфигурация, соответствующая стехиометрической формуле Y_2O_3 . Установлено, что в квазиравновесных условиях в подсистеме кислород-иттрий может возникать нестехиометрия («раскомпенсация»), когда вакансионная подзона оказывается частично заполненной. При этом в вакансионной подрешётке может реализоваться прыжковый механизм токопереноса.

Работа поддержана Российским фондом фундаментальных исследований (проект № 14-07-00582) и базовой темой госзадания № 2629. Расчёты были выполнены на суперкомпьютере ННГУ «Лобачевский».

Ключевые слова: диоксид циркония, кислородные вакансии, допирование иттрием, мемристоры, прыжковая проводимость.

ВВЕДЕНИЕ

В последнее время активно изучаются различные сценарии развития токовых неустойчивостей и необратимых эффектов в наноразмерных плёнках оксидов [Driscoll et al., 2009]. Во многих диоксидах переходных металлов связанные с кислородом дефекты, обычно кислородные вакансии, гораздо более подвижны, чем катионы переходных металлов. Насыщение или обеднение кислородными вакансиями некоторых областей плёнки влияет на валентность катионов переходных металлов, что ведёт к изменению электронной проводимости материала. Поэтому такой механизм резистивного переключения связывают с изменением валентности.

В данной работе исследованы плёнки на основе диоксида циркония (ZrO_2), стабилизированного иттрием. Проводимость ZrO_2 существенным образом определяется кислородными вакансиями, концентрация которых, в свою очередь, зависит от содержания иттрия в диоксиде циркония (степени допирования). Введение иттрия ослабляет связи между ионами кислорода и циркония и облегчает возникновение кислородных вакансий. Для описания проводимости в таких матрицах необходимо знать зонную структуру ZrO_2 , положение уровней, образуемых вакансионными центрами, время релаксации носителей, а также поведением носителей на границе оксид-металл (свойств контактов). В свою очередь, распре-

Горшков Олег Николаевич — доцент, кандидат физико-математических наук, gorshkov@nifti.unn.ru

Дербенева Наталья Викторовна — аспирантка, kurova_natasha@mail.ru

Мухаматчин Камиль Рафаилович — младший научный сотрудник, mykamil@yandex.ru

Сатанин Аркадий Михайлович — профессор, доктор физико-математических наук, sarkady@mail.ru

деление электрических полей в плёнке может существенно влиять на движение и перераспределение вакансий кислорода. В связи с этим могут реализоваться два основных сценария неоднородного распределения вакансий, которые будут определять механизмы переключения между резистивными состояниями в материале. Согласно первому сценарию, вакансии вследствие неустойчивости могут сформировать долгоживущие токовые шнуры (filaments) [Tingkun, 2014], возникающие при напряжении «формовки», которые «запоминают» состояние системы. Второй сценарий может быть связан с формированием плоского слоя с повышенной или пониженной концентрацией вакансий в зависимости от направления поля, например, вблизи контакта (барьера Шоттки).

Чтобы конкретизировать модель электронных состояний и сделать её применимой для расчёта реальных мемристивных элементов, необходимо понять причину возникновения заряженных вакансий, а также изучить зарядовые состояния, в которых они могут находиться в зависимости от концентрации вводимых допирующих примесей. Существует аналогия между электронными свойствами легированных полупроводников и твёрдых электролитов. Например, состояния кислородных вакансий V в ZrO_2 обнаруживают свойства, которые вполне аналогичны состояниям доноров в полупроводниках. Однако в данном случае нельзя полностью использовать эту аналогию для расчёта электронных состояний, поскольку в диоксидах уровни вакансий глубокие, а метод эффективной массы не применим для расчёта таких уровней.

Как уже отмечалось выше, важную роль будет играть допирование диоксида циркония иттрием, присутствие которого может существенно влиять на электронные состояния вакансий. В данной работе показано, что иттрий может играть роль акцептора в рассматриваемых соединениях. Далее для определённости рассматриваются электронные свойства стабилизированного в кубической фазе диоксида циркония. Несмотря на то, что в объёмных материалах стабилизация кубической фазы ZrO_2 имеет место только при концентрациях Y около десятка процентов, в плёнках толщиной $d \approx 10...50$ нм, выращенных на кремниевой подложке, стабильными оказываются слои нелегированного кубического ZrO_2 [Aarik et al., 2002; Holgado et al., 2001] при температуре около $T \approx 180...220$ °C.

В работе представлены детальные расчёты поведения вакансионных состояний в зависимости от степени допирования, что является новым, и обсуждается механизм прыжковой проводимости по вакансионной зоне.

МОДЕЛЬ И МЕТОД РАСЧЁТА

Для расчёта энергии образования электронных состояний вакансий, необходимо рассмотреть большое число атомов, входящих в так называемую суперячейку (ячейка с размерами в несколько постоянных решётки). Например, в настоящей работе в качестве основы выбрана суперячейка $2 \times 2 \times 2$ оксида циркония, содержащая 96 атомов, в которой можно генерировать вакансии путём удаления атомов кислорода и моделировать допирование путём замещения атомов циркония атомами допанта (иттрия). Принципиальная проблема, возникающая на пути увеличения числа частиц, состоит в том, что состояния электронов и ионов в суперячейке характеризуются большим числом степеней свободы, поэтому в общем случае оказывается невозможным прямое численное интегрирование соответствующего уравнения Шредингера. Для решения поставленной задачи мы обратимся к методу функционала плотности (или методу Кона-Шэма — DFT, density functional theory). Отметим, что теория Кона-Шэма лежит в основе современных

расчётов электронных свойств конденсированных систем [Dreizler, Gross, 1990; Hohnberg, Kohn, 1964; Kohn, Sham, 1965; Martin, 2004; Parr, Yang, 1989].

Основное положение метода функционала плотности основывается на том, что важнейшие свойства системы взаимодействующих частиц могут быть выражены с помощью функционала электронной плотности $n(\mathbf{r})$. Плотность электронов — это скалярная функция трёх переменных (точки наблюдения) определяющая, в принципе, всю информацию об основном состоянии и спектре возбуждений. Существование такого функционала для многоэлектронной системы, находящейся при нулевой температуре, впервые было доказано в работах [Hohnberg, Kohn, 1964; Kohn, Sham, 1965]. Практическое значение метод приобрёл только после того, как В. Кона и Л. Шэм предложили подход для вычисления функционала плотности [Kohn, Sham, 1965]. Практический подход Кона-Шэма основывается на двух предположениях: точная электронная плотность основного состояния взаимодействующих электронов может быть заменена на плотность свободных частиц вспомогательной системы; гамильтониан вспомогательной системы выбирается так, чтобы он имел обычную кинетическую энергию и эффективный локальный потенциал, который ответственен за кулоновское взаимодействие, корреляцию и обмен.

Уравнение для орбиталей Кона-Шэма имеет вид:

$$-\frac{1}{2}\nabla^2\psi_i(\mathbf{r})+V_{KS}(\mathbf{r})\psi_i(\mathbf{r})=\varepsilon_i\psi_i(\mathbf{r}), \quad (1)$$

где суммарный потенциал (потенциал Кона-Шэма):

$$V_{KS}(\mathbf{r})=V_{ext}(\mathbf{r})+V_H(\mathbf{r})+V_{xc}(\mathbf{r}) \quad (2)$$

состоит из внешнего потенциала $V_{ext}(\mathbf{r})$, потенциала Хартри

$$V_H(\mathbf{r})=\int\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' \quad (2a)$$

и так называемого обменно-корреляционного потенциала $V_{xc}(\mathbf{r})$ [Becke, Johnson, 2006]. В ряде случаев обменно-корреляционный потенциал можно выразить как функцию локальной электронной плотности, которая связана с волновой функцией обычным выражением:

$$n(\mathbf{r})=\sum_i|\psi_i(\mathbf{r})|^2, \quad (3)$$

где суммирование производится по всем занятым орбиталиям.

Как видно, уравнение Кона-Шэма имеет вид одночастичного уравнения Шредингера для частицы, движущейся в самосогласованном потенциале, структура которого определяется из выражения (2). Все отличие от уравнения Хартри в соотношении (1) содержится в обменно-корреляционном потенциале, а также в интерпретации собственного значения ε_i . Уравнение Кона-Шэма может рассматриваться как формальное обобщение теории Хартри.

Полезная информация об электронных состояниях содержится в локальной плотности состояний, заданной выражением:

$$D(\varepsilon,\mathbf{r})=\sum_i|\psi_i(\mathbf{r})|^2\delta(\varepsilon-\varepsilon_i), \quad (4)$$

а также в парциальных плотностях состояний, определяемых выражением типа (4) (где выбираются только состояние выделенных орбиталей, например, s , p и т. п.).

Программный комплекс Wien2k [Blaha et al., 2001], используемый нами для расчета электронной структуры, реализует решение задачи Кона-Шэма (1) в базе линейно-аризованных присоединённых плоских волн (LAPW) в реальном пространстве с соответствующим заданным шагом по сетке, равным $0,1 \text{ \AA}$. Построение базиса LAPW основывается на разбиении элементарной ячейки на неперекрывающиеся атомные сферы (центр которых совпадает с координатами расположения атомов в ячейке) и область вне этих сфер, где волновая функция представляется в виде разложения по сферическим функциям и плоским волнам соответственно. Для корректного учёта ширины запрещённой зоны был выбран обменно-корреляционный потенциал в форме Беке-Джонсона [Becke, Johnson, 2006].

КИСЛОРОДНЫЕ ВАКАНСИИ В ДИОКСИДЕ ЦИРКОНИЯ

Для начала рассчитаем зонную структуру кристалла диоксида циркония без примесей и дефектов. Полученные результаты расчётов на кластере НИФТИ ННГУ с использованием пакета Wien2k представлены на рис. 1.

Как видно на рис. 1, валентная зона идеального кристалла ZrO_2 формируется p -состояниями кислорода с малой примесью d -состояний циркония, а зона проводимости — d -состояниями циркония [Шапошников и др., 2006]. При этом уровень Ферми E_F расположен на потолке валентной зоны, и кристалл — непроводящий с шириной запрещённой зоны E_g порядка $4,54 \text{ эВ}$. Полученные результаты хорошо коррелируют с известными в литературе, где электронный спектр кристаллов оксида циркония достаточно подробно изучен с применением различных методов расчёта: как в рамках модели Кона-Шэма, с выбором соответствующего обменно-корреляционного потенциала [Шапошников и др., 2006], так и использованием GW-подхода, который учитывает поляризацию решётки [Jiang et al., 2010]. Отметим, что значения для ширины запрещённой зоны ZrO_2 , вычисленные теоретическим путём в целом согласуются с данными, полученными с помощью экспериментальных методов ($E_g \approx 5...6 \text{ эВ}$) [Шапошников и др., 2006; Jiang et al., 2010].

Нарушение трансляционной симметрии решётки путём создания кислородных вакансий приведёт к отщеплению части состояний, образующих зону проводимости и образованию двукратно вырожденных по спину глубоких энергетических уровней в запрещённой зоне кристалла (результаты расчётов представлены на рис. 2). Величина расщепления зависит от концентрации вакансий и меняется примерно от $1,8 \text{ эВ}$ для изолированного дефекта до $1...1,2 \text{ эВ}$ для вакансионных комплексов. Электронный спектр одиночной вакансии не зависит от выбора кислородной подрешётки, однако уже при образовании дивакансии относительное расположение дефектов является важным. Следует отметить, что часть уровней попадут в резонанс с состояниями зоны проводимости (см. рис. 2б и в). Проведение релаксации структуры не оказывает существенного влияния на электронный спектр. Вычисленные значения хорошо согласуются с данными, рассчитанными методом GW, дающим для кубической фазы ZrO_2 локализованный уровень с энергией $2,1 \text{ эВ}$ [Kralik et al., 1997].

Вакансионный центр не может быть ионизован термическими флуктуациями при комнатных температурах или слабыми электрическими полями, поэтому кристалл остаётся диэлектриком, при этом ячейка, содержащая дефект, будет электронейтральной. Таким образом, вакансионные комплексы характеризуются уровнями, лежащими глубоко в запрещённой зоне, и для их описания не применимо приближение эффективной массы.

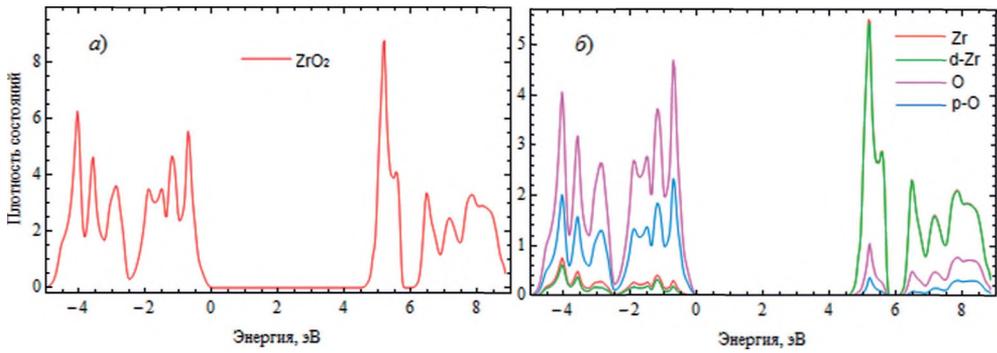


Рис. 1. Плотность состояний идеального кристалла диоксида циркония (а) и парциальные компоненты плотности состояний (б), которая показывает вклад различных орбиталей

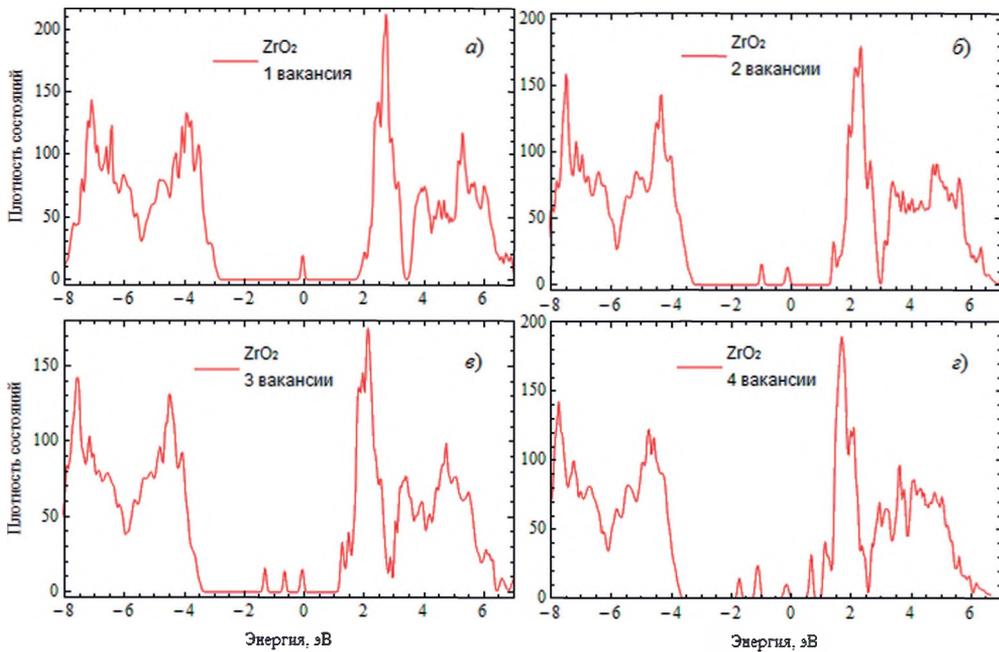


Рис. 2. Положение локализованных уровней в запрещённой кристалла диоксида циркония при генерации в суперячейке одной (а) двух (б) трёх (в) и четырёх (г) кислородных вакансий, расположенных случайным образом. Пунктирной линией показано положение уровня Ферми

Рассмотрим устройство электронных состояний точечного дефекта на примере изолированной одиночной вакансии. Кулоновское взаимодействие двух носителей, оставшихся на оборванных связях после удаления атома кислорода, образует дублет, нижний уровень которого расположен в середине запрещённой зоны, а верхний попадает на дно зоны проводимости. Основной вклад в локализованные состояния будут давать внешние *d*-орбитали атомов циркония и *p*-состояния кислорода из ближайшего окружения вакансии (рис. 3, 4), что полностью согласуется с литературными данными [Králík et al., 1997].

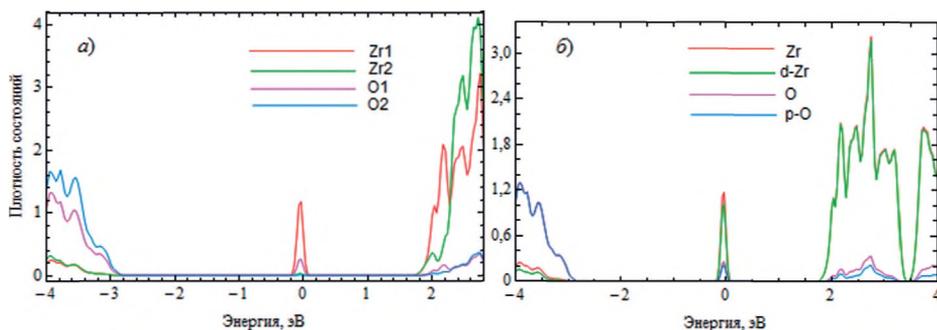


Рис. 3. Сравнительный вклад в плотность состояний атомов циркония и кислорода из первых двух (Zr_1 и O_1) следующих за ними (Zr_2 и O_2) координационных сфер окружения кислородной вакансии (а); полная плотность состояний атомов циркония и кислорода, находящихся в позициях ближайших (Zr) и вторых (O) соседей, а также вклады p -оболочки кислорода и d -оболочки циркония (б)

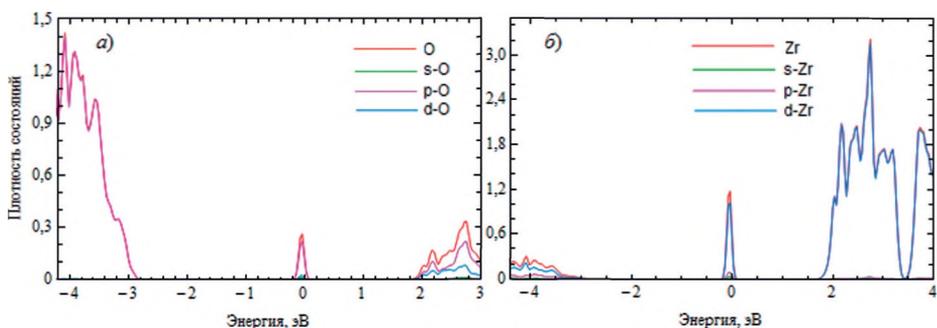


Рис. 4. Полная и частичная плотность состояний, соответствующая s , p и d -орбиталям, атомов кислорода (а) и циркония (б) из ближайшего окружения вакансии

Поскольку, как показывают результаты компьютерного моделирования, основную роль играет ближайшее окружение, проведём аналитическое рассмотрение в рамках модели сильной связи, где электроны, связанные с анионным дефектом будем описывать с помощью волновой функции, составленной из d -орбиталей атомов циркония, стоящих в вершинах тетраэдра, в центре которого расположена вакансия. Выберем начало координат в одной из вершин тетраэдра, тогда оставшиеся катионные узлы будут находиться в точках с радиус-векторами $\mathbf{h}^{(1)} = a(0, 1/2, 1/2)$, $\mathbf{h}^{(2)} = a(1/2, 0, 1/2)$ и $\mathbf{h}^{(3)} = a(1/2, 1/2, 0)$, где a — размер элементарной ячейки кристалла диоксида циркония. Возможны следующие линейные комбинации волновых функций $\varphi_v(\mathbf{r})$, преобразующихся как xy , xz , yz , $x^2 - y^2$ и $3z^2 - r^2$, ортогональные друг другу $\langle \psi_{\nu}(\mathbf{r}) | \psi_{\nu'}(\mathbf{r}) \rangle = \delta_{\nu\nu'}$:

$$\left. \begin{aligned} \psi_{\nu_1}(\mathbf{r}) &= A_{\nu_1} \left\{ \varphi_{\nu}(\mathbf{r}) + \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(1)}) - \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(2)}) - \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(3)}) \right\} / 2, \\ \psi_{\nu_2}(\mathbf{r}) &= A_{\nu_2} \left\{ \varphi_{\nu}(\mathbf{r}) - \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(1)}) + \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(2)}) - \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(3)}) \right\} / 2, \\ \psi_{\nu_3}(\mathbf{r}) &= A_{\nu_3} \left\{ \varphi_{\nu}(\mathbf{r}) - \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(1)}) - \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(2)}) + \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(3)}) \right\} / 2, \\ \psi_{\nu_4}(\mathbf{r}) &= A_{\nu_4} \left\{ \varphi_{\nu}(\mathbf{r}) + \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(1)}) + \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(2)}) + \varphi_{\nu}(\mathbf{r} - \mathbf{h}^{(3)}) \right\} / 2, \end{aligned} \right\} \quad (5)$$

здесь $A_{\nu l}$ — нормировочные константы. Легко видеть, что преобразование типа $\mathbf{r} \rightarrow -\mathbf{r} + \mathbf{h}^{(j)}$ либо оставляет знак $\psi_{\nu l}(\mathbf{r})$, либо меняет его на противоположный. С учётом данного обстоятельства из волновых функций (5) можно построить ортогонализированный базис функций $\Psi_i(\mathbf{r})$, удовлетворяющих условию нормировки $\langle \Psi_i(\mathbf{r}) | \Psi_j(\mathbf{r}) \rangle = \delta_{ij}$, в котором гамильтониан электрона, движущегося в поле кристаллической решётки ZrO_2 (с вакансией), примет диагональный вид. Опираясь на этот базис, пользуясь процедурой, описанной в литературе [Стоунхэм, 1978], можно найти электронные состояния вакансии.

Вначале запишем два набора функций $\Psi_{n\pm}(\mathbf{r})$, обозначаемых индексами «+» и «-», которым соответствуют два трёхкратно вырожденных уровня $E_{s\pm}$, причём Ψ_i в каждом из наборов преобразуются друг через друга с сохранением знака:

$$\left. \begin{aligned} \Psi_{1,\pm}(\mathbf{r}) &= B_{\pm} \left\{ \lambda \left[\psi_{xy,1}(\mathbf{r}) + \psi_{xz,4}(\mathbf{r}) + \psi_{yz,3}(\mathbf{r}) \right] / 3^{1/2} \pm \left[\psi_{3z^2-r^2,2}(\mathbf{r}) + 3^{1/2} \psi_{x^2-y^2,2}(\mathbf{r}) \right] / 2 \right\}, \\ \Psi_{2,\pm}(\mathbf{r}) &= B_{\pm} \left\{ \lambda \left[\psi_{xy,2}(\mathbf{r}) + \psi_{xz,3}(\mathbf{r}) + \psi_{yz,4}(\mathbf{r}) \right] / 3^{1/2} \pm \left[\psi_{3z^2-r^2,1}(\mathbf{r}) - 3^{1/2} \psi_{x^2-y^2,1}(\mathbf{r}) \right] / 2 \right\}, \\ \Psi_{3,\pm}(\mathbf{r}) &= B_{\pm} \left\{ \lambda \left[\psi_{xy,4}(\mathbf{r}) + \psi_{xz,1}(\mathbf{r}) + \psi_{yz,2}(\mathbf{r}) \right] / 3^{1/2} \pm \psi_{3z^2-r^2,3}(\mathbf{r}) \right\}, \end{aligned} \right\} \quad (6)$$

B_{\pm} — нормировочные константы, параметр λ найдём из требования:

$$\frac{\lambda^2}{3} \int \left[\psi_{xy,1}(\mathbf{r}) + \psi_{xz,4}(\mathbf{r}) + \psi_{yz,3}(\mathbf{r}) \right]^2 dV = 1.$$

Отметим, что вероятность обнаружения электрона в некоторой точке кристалла, описываемая системой функций (6), достигает своего минимума (нуля) на кислородном дефекте и максимума за пределами координационного многогранника (тетраэдра).

Другие возможные линейные комбинации атомных функций $\psi_{\mu}(\mathbf{r})$, ортогональные $\Psi_i(\mathbf{r})$, описывают ситуацию, при которой минимум электронной плотности (равный нулю) также находится в месте нахождения вакансии, а максимум внутри тетраэдрического объёма, образованного атомами циркония из ближайшего окружения, при этом $\Psi'_i(\mathbf{r})$ имеют узлы в вершинах тетраэдра (атомах Zr):

$$\left. \begin{aligned} \Psi'_{1,\pm}(\mathbf{r}) &= C_{\pm} \left\{ \frac{\beta}{\sqrt{2}} \left[\psi_{xy,1}(\mathbf{r}) - \psi_{yz,3}(\mathbf{r}) \right] \pm \left[3^{1/2} \psi_{3z^2-r^2,2}(\mathbf{r}) - \psi_{x^2-y^2,2}(\mathbf{r}) \right] / 2 \right\}, \\ \Psi'_{2,\pm}(\mathbf{r}) &= C_{\pm} \left\{ \frac{\beta}{\sqrt{2}} \left[\psi_{xz,3}(\mathbf{r}) - \psi_{xy,2}(\mathbf{r}) \right] \pm \left[3^{1/2} \psi_{3z^2-r^2,1}(\mathbf{r}) + \psi_{x^2-y^2,1}(\mathbf{r}) \right] / 2 \right\}, \\ \Psi'_{3,\pm}(\mathbf{r}) &= C_{\pm} \left\{ \frac{\beta}{\sqrt{2}} \left[\psi_{yz,2}(\mathbf{r}) - \psi_{xz,1}(\mathbf{r}) \right] \pm \psi_{x^2-y^2,3}(\mathbf{r}) \right\}. \end{aligned} \right\} \quad (7)$$

C_{\pm} — нормировочные константы, а β удовлетворяет условию:

$$\frac{\beta^2}{2} \int \left[\psi_{xy,1}(\mathbf{r}) - \psi_{yz,3}(\mathbf{r}) \right]^2 dV = 1.$$

Волновым функциям (7) также отвечают два трёхкратно вырожденных энергетических уровня $E_{a\pm}$. Отметим, что при перестановке осей $\Psi'_{i,\pm}(\mathbf{r})$ преобразуются друг через друга с изменением знака.

Ортогональные (8) и (9) функции

$$\Psi_{1,\pm}''(\mathbf{r}) = L \left\{ \frac{\gamma}{\sqrt{2}} [\psi_{xz,2}(\mathbf{r}) - \psi_{yz,1}(\mathbf{r})] \pm \psi_{x^2-y^2,4}(\mathbf{r}) \right\}, \quad (8)$$

$$\Psi_2''(\mathbf{r}) = \psi_{3z^2-r^2,4}(\mathbf{r}), \quad (9)$$

где γ подчиняется условию:

$$\frac{\gamma^2}{2} \int [\psi_{xy,1}(\mathbf{r}) - \psi_{yz,3}(\mathbf{r})]^2 dV = 1,$$

а L определяется из требования нормировки, описывают ситуации, при которых электронная плотность на кислородной вакансии и окружающих её атомах циркония равна нулю и достигает максимума в межъядерном пространстве.

И, наконец, волновая функция, ортогональная (6)–(9) и обладающая симметрией кристаллической решётки (относительно точки расположения дефекта), и описывающая основное состояние локализованного на вакансии электрона имеет вид:

$$\Psi_v(\mathbf{r}) = D [\psi_{xy,3}(\mathbf{r}) + \psi_{xz,2}(\mathbf{r}) + \psi_{yz,1}(\mathbf{r})], \quad (10)$$

здесь D — нормировочная константа.

Отметим, что согласно результатам численного моделирования оказалось, что в плотность состояний дефекта дадут вклад только $t2g$ орбитали валентной оболочки циркония. Данный факт полностью учитывается видом функции $\Psi_v(\mathbf{r})$. Отношение вероятностей нахождения d -электрона на вакансии A_v и атоме циркония (в одной из вершин тетраэдра) A_{Zr} определяется отношением радиуса координационной сферы d_0 , описанной вокруг вакансии к длине ребра вписанного в неё тетраэдра h_t :

$$A_v = \left(\frac{2}{3} \right)^2 \frac{R^2 \left(a\sqrt{3}/4 \right)}{R^2 \left(a\sqrt{2} \right)} A_{Zr},$$

$R(\mathbf{r})$ — радиальная часть волновой функции, удовлетворяющая требованию:

$$\int r^2 R^2(\mathbf{r}) dV = 1.$$

При достаточно быстром затухании $R(\mathbf{r})$ $A_v \gg A_{Zr}$, заряд в основном распределён в области дефекта.

Увеличение концентрации вакансий (образование дивакансии) приведёт к расщеплению энергетического уровня основного состояния на два подуровня, при этом качественная картина не изменится.

ВЛИЯНИЕ ЛЕГИРОВАНИЯ ИТТЕРИЕМ НА ЗАПОЛНЕНИЕ ВАКАНСИОННЫХ УРОВНЕЙ

Рассмотрим теперь эффект от введения атомов иттрия в решётку диоксида циркония.

Поскольку электронная конфигурация внешней оболочки атома Y отличается от оболочки Zr лишь заполнением d -орбитали, а все энергии ионизации внешних электронов близки [Зайнуллина и др., 2000], при замещении атомов циркония иттрием в запрещённой зоне не возникнет дополнительных уровней (рис. 5).

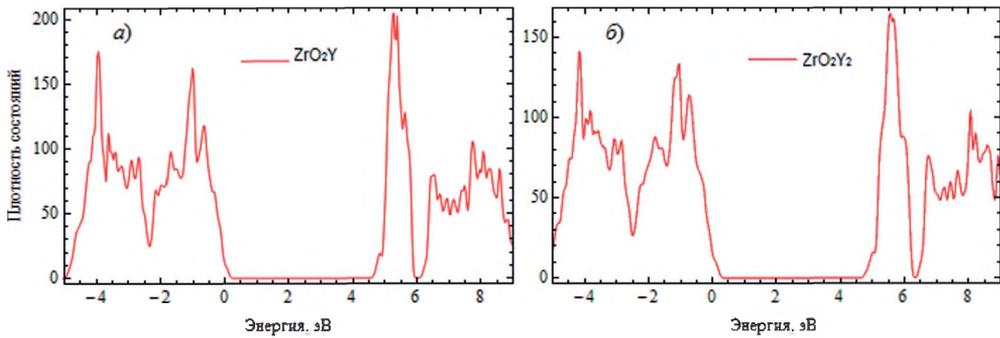


Рис. 5. Плотность состояний оксида циркония, допированного иттрием в концентрации 1 атом Y на суперячейку (а), 2 атома Y на суперячейку (б)

Однако, имея меньшую валентность, иттрий может проявлять акцепторные свойства. В результате при образовании вакансии один электрон из состояния, локализованного на дефекте, перейдёт в d -состояние атома допанта, достраивая вакансию до электронной конфигурации циркония и участвуя в образовании связей иттрия с кислородным окружением. Отметим, что в силу большей локализации электронной плотности вблизи иттрия, его вклад в плотность электронных состояний вакансии значительно меньше, чем у циркония, находящегося в том же окружении.

Как следствие результата эффекта компенсации, уровень Ферми понижается, вакансионная зона частично опустошается и становится возможна прыжковая проводимость по локализованным состояниям. Часть вакансий при этом переходит в заряженное состояние. При увеличении концентрации иттрия, когда на одну вакансию будет приходиться по два атома допанта, наступит эффект полной компенсации. Уровень Ферми окажется на потолке валентной зоны и плёнка вновь перейдёт в диэлектрическую фазу.

ЗАКЛЮЧЕНИЕ

В работе показано, что иттрий, замещающий атомы циркония в ZrO_2 при допировании играет роль акцептора и, тем самым, частично опустошает локализованные вакансионные состояния, пиннигуя на них уровень Ферми. Если же уровень Ферми попадает в область пика плотности состояний, то может стать возможным прыжковый перенос по вакансионной подзоне. При этом с увеличением концентрации вакансий происходит размытие локализованных состояний в вакансионную зону, а степень заполнения определяется концентрацией атомов допанта, с переходами из диэлектрической фазы (недопированный кристалл) в проводящее состояние и возвращением в состояние изолятора при полной компенсации, когда концентрация вакансий равна двум концентрациям иттрия.

Отметим, что образование регулярной структуры при выстраивании вакансий вдоль одного из направлений типа (100) показывает, что их состояния могут привести к локальной модификации свойств кристалла с образованием металлического шнура, электронные состояния которого описываются появлением соответствующей минизоны (α -зоны) в запрещённой зоне диэлектрика.

ЛИТЕРАТУРА

- [Зайнуллина и др., 2000] *Зайнуллина В. М., Жуков В. П., Жуковский В. М., Медведева Н. И.* Электронная структура и характеристики ионной проводимости в диоксиде циркония, стабилизированном примесями кальция и иттрия // Журн. структурной химии. 2000. Т. 41. № 2. С. 229–239.
- [Стоунхэм, 1978] *Стоунхэм А. М.* Теория дефектов в твёрдых телах. В 2-х т. М.: Мир. 1978. 928 с.
- [Шапошников и др., 2006] *Шапошников А. В., Гриценко Д. В., Петренко И. П., Пчеляков О. П., Гриценко В. А., Эренбург С. Б., Бауск Н. В., Бадалян А. М., Шубин Ю. В., Смирнова Т. П., Вонг Х., Ким Ч. В.* Атомная и электронная структура ZrO_2 // ЖЭТФ. 2006. Т. 129. № 5. С. 914–925.
- [Aarik et al., 2002] *Aarik J., Aidla A., Mändar H., Uustare T., Sammelselg V.* Growth kinetics and structure formation of ZrO_2 thin films in chloride-based atomic layer deposition process // Thin Solid Films. 2002. V. 408. No. 1-2, 3. P. 97–103.
- [Becke, Johnson, 2006] *Becke A. D., Johnson E. R.* A simple effective potential for exchange // J. Chem. Phys. 2006. V. 124. P. 221101.
- [Blaha et al., 2001] *Blaha P., Schwarz K., Madsen G. K. H., Kvasnicka D., Luitz J.* WIEN2k, An Augmented Plane Wave + Local Orbitals Program for Calculating Crystal Properties / Karlheinz Schwarz, Techn. Universitat Wien, Austria. 2001. 258 p.
- [Dreizler, Gross, 1990] *Dreizler R. M., Gross E. K. V.* Density Functional Theory. Springer-Verlag, Berlin, Heidelberg. 1990. 676 p.
- [Driscoll et al., 2009] *Driscoll T., Kim H.-T., Chae B.-G., Kim B.-J., Lee Y.-W., Jokerst N. M., Palit S., Smith D. R., Di Ventura M., Basov D. N.* Memory metamaterials // Science. 2009. V. 325. P. 1518–1521.
- [Hohnberg, Kohn, 1964] *Hohnberg P., Kohn W.* Inhomogeneous Electron Gas // Phys. Rev. A. 1964. V. 136. P. B864–B871.
- [Holgado et al., 2001] *Holgado J. P., Espinos J. P., Yubero F., Justo A., Ocana M., Benitez J., Gonzalez-Elip A. R.* Ar stabilisation of the cubic tetragonal phases of ZrO in thin films prepared by ion beam induced chemical vapour deposition // Thin Solid Films. 2001. V. 389. P. 34–42.
- [Jiang et al., 2010] *Jiang H., Gomez-Abal R. I., Rinke P., Scheffler M.* Electronic band structure of zirconia and hafnia polymorphs from the GW perspective // Phys. Rev. B. 2010. V. 81. P. 085119-1–085119-9.
- [Kohn, Sham, 1965] *Kohn W., Sham L. J.* Self-Consistent Equations Including Exchange and Correlation Effects // Phys. Rev. 1965. V. 140. P. A1133–A1138.
- [Králík et al., 1997] *Králík B., Chang E. K., Louie S. G.* Structural properties and quasiparticle band structure of zirconia // Phys. Rev. B. 1997. V. 57. P. 7027–7036.
- [Martin, 2004] *Martin R. M.* Electronic Structure: Basic Theory and Practical Methods. Cambridge: Cambridge University Press, 2004. 648 p.
- [Parr, Yang, 1989] *Parr R. G., Yang W.* Density Functional Theory of Atoms and Molecules. Oxford: Oxford University Press, 1989. 333 p.
- [Tingkun, 2014] *Tingkun Gu.* Metallic filament formation by aligned oxygen vacancies in ZnO-based resistive switches // J. Appl. Phys. 2014. V. 115. P. 203707-1–203707-8.

NUMERICAL SIMULATION OF ELECTRON STATES IN THE ZIRCONIUM DIOXIDE WITH OXYGEN VACANCIES AND YTTRIUM DOPANT

O. N. Gorshkov, N. V. Derbeneva, K. R. Muhamatchin, A. M. Satanin

Lobachevsky State University of Nizhniy Novgorod (UNN)

Electronic states in thin films based on zirconium dioxide (ZrO_2) stabilized with yttrium (Y) were studied. Calculations of the electronic states in doped and undoped zirconium dioxide were performed within the Kohn-Sham method. Restructurization of the vacancy states was investigated depending on the concentration of yttrium (the degree of doping). Under certain conditions of the equilibrium position a configuration of the stoichiometric formula Y_2O_3 in the subsystem oxygen-yttrium stabilized. It was found that in quasi-equilibrium conditions in the oxygen-yttrium subsystem non-stoichiometry (“decompensation”) can occur when vacancy subband is partially filled. At the same time the hopping mechanism of charge transport can be realized in the vacancy sublattice.

Keywords: zirconium dioxide, oxygen vacancies, doping with yttrium, memristors, hopping conductivity

Gorshkov Oleg Nicolaevich — assistant professor, PhD, gorshkov@nifti.unn.ru

Derbeneva Natalia Victorovna — postgraduate student, kurova_natasha@mail.ru

Muhamatchin Kamil Rafailovich — junior researcher, mykamil@yandex.ru

Satanin Arkady Mikhailovich — professor, doctor of physical and mathematical sciences, sarkady@mail.ru

ВЛИЯНИЕ ФОСФОРА НА ОЖЕ-РЕКОМБИНАЦИЮ В НАНОКРИСТАЛЛАХ КРЕМНИЯ

Н. В. Дербенева

Нижегородский государственный университет им. Н. И. Лобачевского (ННГУ)

В работе представлены результаты расчётов электронной структуры кремниевых нанокристаллов, легированных фосфором, и скоростей Оже-рекомбинации. Показано, что наличие нескольких атомов фосфора приводит к более сильному уменьшению ширины оптической щели и к росту скорости Оже-процесса по сравнению со случаем одного атома.

Работа поддержана Российским фондом фундаментальных исследований (проекты № 14-07-00582-а, 14-02-31637-мол_а).

Ключевые слова: кремний, нанокристалл, теория функционала плотности, оже-процесс.

ВВЕДЕНИЕ

Изучение нанокристаллического кремния в последние годы стало весьма актуальным в связи с перспективами его применения в полупроводниковой микроэлектронике, например, при создании светоизлучающих приборов. Это связано, в первую очередь, с обнаружением довольно интенсивной фотолюминесценции от структур на его основе. Тем не менее, как показывают результаты расчётов [Hybertsen, 1994] и экспериментов [Kanemitsu, 1996], характерные значения скоростей излучательной рекомбинации в кремниевых кристаллитах оказываются на порядки меньше скоростей конкурирующих безызлучательных процессов, в частности, Оже-рекомбинации.

Данный механизм безызлучательной релаксации электронных возбуждений наиболее вероятен в случае, когда суммарное число носителей заряда в обеих зонах (проводимости и валентной) становится больше двух. Согласно результатам расчётов, проводившихся ранее на основе приближения огибающей [Mahdouani et al., 2009], методов сильной связи [Deleue, Lannoo, 2004] и эмпирического псевдопотенциала [Sevik, Bukutay, 2008], характерные значения скорости Оже-рекомбинации варьируются в пределах от 10^6 до 10^{11} с^{-1} для случая «чистых» нанокристаллов, в то время как скорости излучательной рекомбинации принимают значения от 10^2 до 10^5 с^{-1} . В связи с этим возникает необходимость модификации электронной структуры кремниевых нанокластеров с целью подавления подобных «паразитных» процессов.

Одним из методов воздействия на электронные состояния в нанокристаллах является допирование их различными примесями. В настоящее время достаточно большое число работ посвящено изучению влияния типа легирующей примеси на интенсивность фотолюминесценции от материалов на основе нанокристаллов кремния. При этом довольно много внимания уделяется роли «изокорных» атомов, в частности, фосфору, практически не вносящему изменений в геометрию кристаллической структуры нанокластеров. Как показывают исследования [Belyakov et al., 2009], легирование фосфором способствует заметному усилению

(до 5 раз) интенсивности излучательной рекомбинации, что является следствием модификации электронной структуры нанокристаллов полем примесных центров. Однако для получения полной картины фотолуминесценции необходимо также учитывать влияние фосфора и на скорости безызлучательных процессов. В связи с этим, в данной работе рассчитывается электронная структура и скорости Оже-рекомбинации для случая *eeh*-процесса в кремниевых кристаллитах, легированных фосфором. Расчёт проводился как для одного донора в зависимости от его положения внутри нанокластера, так и для нескольких доноров, концентрация n_D которых достигала $\sim 2\%$.

ЭЛЕКТРОННАЯ СТРУКТУРА КРЕМНИЕВЫХ НАНОКЛАСТЕРОВ, ЛЕГИРОВАННЫХ ФОСФОРОМ

Основой для исследования скоростей Оже-процесса служит знание особенностей электронной структуры кристаллитов в присутствии донора, для расчёта которой было проведено моделирование структуры нанокристаллов размерами от одного до двух нанометров. Формирование нанокластеров происходило последовательным заполнением координационных сфер, центрированных относительно узла, соответствующего положению атома в объёмном кремнии. При исследовании электронных свойств нанокристаллического кремния, как правило, рассматривается два типа окружения: либо кристаллит помещается в диэлектрическую матрицу (к примеру, SiO_2), либо оборванные связи на его границе насыщаются водородом, а сам он при этом находится в вакууме. В обоих случаях создаётся так называемый потенциал конфайнмента, препятствующий выходу электронов за пределы нанокристалла, а также появлению внутри запрещённой зоны состояний, связанных с наличием нескомпенсированных оборванных связей. В данной работе была рассмотрена модель нанокристалла с границей, полностью пассивированной водородом.

Для изучения влияния атома фосфора на электронные свойства нанокристалла кремния были смоделированы следующие структуры размерами 1...2 нм: «чистые» нанокластеры, кристаллиты с одним атомом фосфора, замещающим кремний в одной из координационных сфер, а также случай нескольких доноров ($n_D \approx 2\%$), расположенных симметрично относительно центра нанокристалла.

Релаксация кристаллической структуры нанокластера, выполненная в рамках подхода, предложенного в работе Р. Фейнмана [Feynman, 1939], позволила практически полностью (до 0,2 эВ/Å) уменьшить величину межатомных сил, возникающих в связи с формированием нанокристаллов из объёмного кремния.

Расчёт электронных и дырочных состояний нанокластеров был проведён в рамках стационарной теории функционала плотности (density functional theory, DFT) [Hohenberg, Kohn, 1964; Kohn, Sham, 1965] на основе самосогласованного расчёта электронной плотности и полной энергии системы. Применение такого подхода позволило естественным образом преодолеть проблему учёта взаимодействия электронов с примесным центром и между собой. Для расчётов использовался программный комплекс Octopus [Castro et al., 2006], который реализует решение задачи Кона-Шема в реальном пространстве. Для определения наилучшего шага по координате Δx был проведён расчёт полной энергии заданной системы при разных значениях Δx до достижения условия сходимости, равного 0,01 эВ. Эта величина обеспечивает достаточно хорошее описание полной энергии системы и оптимальное время расчёта. В результате обнаружилось, что для нанокристаллов размерами 1–2 нм значение Δx можно выбрать равным 0,2 Å, что и было

сделано. В качестве граничного условия применялось обращение в нуль волновой функции на поверхности кристаллита. Сходимость самосогласованного цикла определялась по разнице между значениями полных энергий, которые соответствовали двум последовательным этапам цикла исследуемой системы. Эта величина была выбрана равной 10^{-5} эВ. Для обменно-корреляционного слагаемого использовалось обобщённо-градиентное приближение в модели PBE [Perdew et al., 1996].

Как известно, стационарная теория Кона-Шема приводит к ошибкам в определении ширины запрещённой зоны исследуемых материалов, а также энергий возбуждённых состояний. В связи с этим был применен метод нестационарной теории функционала плотности в приближении Касиды [Casida et al., 1998], который позволяет в ходе самосогласованного расчёта уточнить энергетический спектр зоны проводимости. Данный подход разработан в рамках теории линейного отклика и может быть применён, в силу своей формулировки, лишь к системам, имеющим конечные размеры. В его основе лежит определение коэффициента поглощения $\alpha(\omega)$, который можно представить в виде суперпозиции вкладов от всех состояний, причём в знаменателе каждого слагаемого стоит величина, отвечающая разнице энергий между соответствующими уровнями:

$$\alpha(\omega) = \sum_I \frac{f_I}{\omega_I^2 - \omega^2}, \quad (1)$$

где f_I — силы осцилляторов соответствующих переходов. Такое описание, при условии, что значения энергий электронных уровней уже были получены ранее в рамках теории DFT, сводится к поиску собственных векторов и собственных значений, задающих частоты оптических переходов в системе, а значит, и определяющие ширину оптической щели.

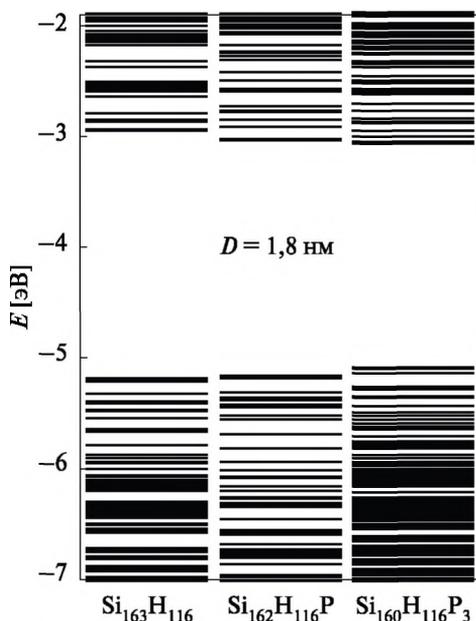


Рис. 1. Электронная структура кремниевого нанокристалла диаметром 1,8 нм: нелегированный, с одним атомом фосфора, а также с тремя атомами фосфора, расположенными симметрично относительно его центра

На рис. 1 представлены результаты расчётов электронной структуры кремниевых нанокристаллов диаметром 1,8 нм с различной концентрацией атома фосфора.

Использование метода Касиды для уточнения ширины оптической щели приводит к поправкам, составляющим 0,23, 0,26 и 0,27 эВ соответственно. Её зависимость от размера объясняется тем, что с изменением количества примеси в нанокристалле происходит и существенное изменение обменно-корреляционного потенциала, который, фактически, и определяет величину данной поправки. Как видно на рис. 1, при легировании нанокристалла фосфором, донорный уровень отщеплён вглубь запрещённой зоны. Величина расщепления составляет 0,15 эВ для одного примесного атома и 0,11 эВ для трёх, почти на порядок превышая расщепление в объёмном кремнии. Это говорит о сильном влиянии квантового конфайнмента на электронную структуру уровней нанокластера.

ОЖЕ-ПРОЦЕСС

На основе полученных спектров был проведён расчёт скорости Оже-процесса для нанокристаллов размерами 1...2 нм в рамках первого порядка теории возмущений согласно «золотому» правилу Ферми:

$$\frac{1}{\tau_{if}} = \frac{2\pi}{\hbar} |U_{if}|^2 \delta(E_i - E_f). \quad (2)$$

Как известно, в силу дискретности спектра, Оже-процесс возможен только в том случае, когда энергия межзонного перехода E_g (рис. 2), передаваемая одним электроном другому, совпадёт с энергией соответствующего внутрizonного перехода ΔE_k в зоне проводимости.

В дальнейшем, при расчёте скоростей Оже-рекомбинации для *eeh*-процесса предполагалось, что два возбуждённых электрона в зоне проводимости находятся на нижнем уровне энергии, и дырка в валентной зоне также находится в своём основном состоянии — т.е. занимает верхний уровень. В результате кулоновского взаимодействия электронов один из них переходит в основное состояние валентной зоны, а другой — на более высокий уровень по энергии в зоне проводимости. Скорость перехода определяется матричным элементом кулоновского взаимодействия, который имеет две составляющие: дальнедействующую $V(\mathbf{r}_1, \mathbf{r}_2)$ и короткодействующую $W(\mathbf{r}_1, \mathbf{r}_2)$. Дальнедействующая составляющая, являющаяся водородоподобной в объёмном полупроводнике, в нанокристалле модифицируется за счёт возникновения поляризационных зарядов на границе нанокристалла. Эту часть потенциальной энергии можно представить в следующем виде [Belyakov et al., 2007]:

$$V(r_1, r_2) = \frac{e^2}{\epsilon_s |r_1 - r_2|} + \frac{e^2(\epsilon_s - \epsilon_d)}{\epsilon_s R} \sum_{n=0}^{\infty} \frac{(n+1)P_n(\cos \theta)r_1^n r_2^n}{(n\epsilon_s + (n+1)\epsilon_d)R^{2n}}. \quad (3)$$

Короткодействующая часть потенциала была взята согласно статье [Walter, Cohen, 1970]:

$$W(\mathbf{r}, \mathbf{h}) = -\frac{e^2}{|\mathbf{r} - \mathbf{h}|} \left| A \exp(-\alpha |\mathbf{r} - \mathbf{h}|) + \left| 1 - A - \frac{1}{\epsilon_s} \right| \exp(-\beta |\mathbf{r} - \mathbf{h}|) \right|, \quad (4)$$

где $A = 1,142$; $\alpha = 0,82/a_b$; a_b — боровский радиус; ϵ_s и ϵ_d — диэлектрические проницаемости кремния и вакуума, соответственно.

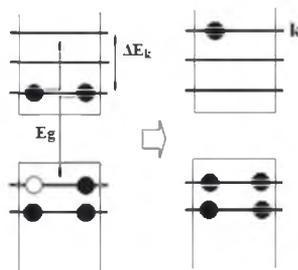


Рис. 2. Механизм Оже-рекомбинации eeh-типа



Рис. 3. Зависимость скорости Оже-процесса от размеров нанокристалла, а также от концентрации фосфора в нём. Каждому из диаметров соответствуют следующие конфигурации нанокластеров: 1,0 нм — $\text{Si}_{35}\text{H}_{36}$; 1,2 нм — $\text{Si}_{59}\text{H}_{60}$; 1,4 нм — $\text{Si}_{87}\text{H}_{76}$; 1,6 нм — $\text{Si}_{123}\text{H}_{100}$; 1,8 нм — $\text{Si}_{147}\text{H}_{100}$; 2,0 нм — $\text{Si}_{175}\text{H}_{116}$

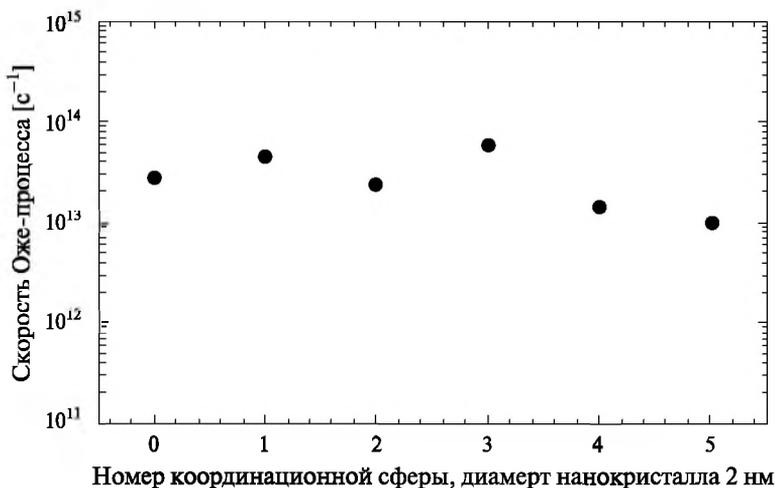


Рис. 4. Зависимость скорости Оже-рекомбинации от положения атома фосфора внутри нанокристалла размером 2,0 нм

На рис. 3 представлены результаты расчётов скоростей Оже-процесса в зависимости от размера нанокристалла для разных типов легирования. Наличие атома фосфора сильно сказывается (происходит усиление на 2–3 порядка) на скорости Оже-рекомбинации, а присутствие нескольких доноров приводит к небольшому увеличению соответствующих значений по сравнению со скоростями в случае одного донора. С увеличением размера нанокристалла скорость Оже-процесса уменьшается, что хорошо согласуется с данными, представленными в работе С. Делерю [Deleue, Lannoo, 2004].

Зависимость скоростей Оже-рекомбинации от позиции атома фосфора внутри нанокристалла, определяемой номером координационной сферы (рис. 4), оказалась довольно слабой, что позволяет судить о характерной скорости изучаемого процесса, не ссылаясь при этом на конкретную позицию примеси в кристаллите. Стоит отметить, что и при легировании нанокристаллов фосфором характерные скорости Оже-процесса остаются все также много выше соответствующих скоростей излучательной рекомбинации, что говорит о необходимости поиска другого пути подавления Оже-рекомбинации.

По сравнению с литературными данными [Deleue, Lannoo, 2004; Mahdouani et al., 2009; Sevik, Bukutay, 2008] рассчитанные значения скоростей Оже-процесса в случае нелегированных нанокристаллов оказались на 1–2 порядка завышенными. Это различие в первую очередь связано с использованием в данной работе нанокристаллов небольших размеров (всего 1...2 нм), в которых сильное влияние на строение электронного спектра оказывает граница, в то время как для нанокластеров больших диаметров (3...4 нм) кристаллическая структура становится близкой к объёмному кремнию. Ещё один фактор, определяющий значение скоростей Оже-процесса, — вид потенциала электрон-электронного взаимодействия. Как показали расчёты, наибольший вклад в скорость Оже-рекомбинации даёт слабое, связанное с короткодействующим потенциалом [Walter, Cohen, 1970]. При этом, как было отмечено ранее этот потенциал должен существенно отличаться от «объёмного» для нанокристаллов малых размеров.

Расчёты были выполнены на суперкомпьютере «Лобачевский» (ННГУ, Нижний Новгород).

ЗАКЛЮЧЕНИЕ

В данной работе был проведён расчёт электронной структуры нанокристаллов кремния с размерами 1...2 нм и показано, что наличие фосфора приводит к отщеплению синглетного донорного уровня вглубь запрещённой зоны на величину, которая оказывается на порядок больше соответствующего расщепления в объёмном кремнии, что объясняется сильным влиянием квантового конфайнмента на электронные свойства нанокристаллов. Значения скоростей Оже-процесса довольно резко возрастают (на 2–3 порядка) в присутствии фосфора, причём усиление происходит соответственно увеличению концентрации легирующей примеси.

ЛИТЕРАТУРА

- [Belyakov et al., 2007] *Belyakov V.A., Burdov V.A.* Valley-orbit splitting in doped nanocrystalline silicon: *k* p calculations // *Phys. Rev. B* 2007. V. 76. P. 045335-1–045335-12.
- [Belyakov et al., 2009] *Belyakov V.A., Belov A. I., Mikhaylov A. N.* et al. Improvement of the photon generation efficiency in phosphorus-doped silicon nanocrystals: Γ -X mixing

- of the confined electron states // *J. Phys.: Condens. Matter*. 2009. V. 21. P. 045803-1–045803-5.
- [Casida et al., 1998] *Casida M. E., Jamorski C., Casida K. C.* et al. Molecular excitation energies to high-lying bound states from time-dependent density-functional response theory: Characterization and correction of the time-dependent local density approximation ionization threshold // *J. Chem. Phys.* 1998. V. 108. P. 4439–4449.
- [Castro et al., 2006] *Castro A., Appel H., Olivera M.* et al. Octopus: a tool for the application of time-dependent density functional theory // *Physica Status Solidi B*. 2006. V. 243. P. 2465–2488.
- [Delerue, Lannoo, 2004] *Delerue C. Lannoo M.* Nanostructures. Theory and Modeling. Springer Verlag, Berlin-Heidelberg, 2004.
- [Feynman, 1939] *Feynman R. P.* Forces in Molecules // *Phys. Rev.* 1939. V. 56. P. 340–343.
- [Hohenberg, Kohn, 1964] *Hohenberg P., Kohn W.* Inhomogeneous Electron Gas // *Phys. Rev* 1964. V. 136. P. B864–B871.
- [Hybertsen, 1994] *Hybertsen M. S.* Absorption and emission of light in nanoscale silicon structures // *Phys. Rev. Lett.* 1994. V. 72. P. 1514–1517.
- [Kanemitsu, 1996] *Kanemitsu Y.* Photoluminescence spectrum and dynamics in oxidized silicon nanocrystals: A nanoscopic disorder system // *Phys. Rev. B* 1996. V. 53. P. 13515–13520.
- [Kohn, Sham, 1965] *Kohn W., Sham L. J.* Self-Consistent Equations Including Exchange and Correlation Effects // *Phys. Rev.* 1965. V. 140. P. A1133–A1138.
- [Mahdouani et al., 2009] *Mahdouani M., Bourguiga R., Jaziri S.* et al. Investigation of Auger recombination in Ge and Si nanocrystals embedded in SiO₂ matrix // *Physica E*. 2009. V. 42. P. 57–62.
- [Perdew et al., 1996] *Perdew J. P., Burke K., Ernzerhof M.* Generalized Gradient Approximation Made Simple // *Phys. Rev. Lett.* 1996. V. 77. P. 3865–3868.
- [Sevik, Bukutay, 2008] *Sevik C. Bukutay C.* Auger recombination and carrier multiplication in embedded silicon and germanium nanocrystals // *Phys. Rev. B*. 2008. V. 77. P. 125414-1–125414-4.
- [Walter, Cohen, 1970] *Walter J. P., Cohen M. L.* Wave-Vector-Dependent Dielectric Function for Si, Ge, GaAs, and ZnSe // *Phys. Rev. B*. 1970. V. 2. P. 1821–1826.

EFFECT OF PHOSPHORUS DOPING ON AUGER RECOMBINATION RATES IN SILICON NANOCRYSTALS

N. V. Derbeneva

Lobachevsky State University of Nizhniy Novgorod (UNN)

Results of calculations of the electronic structure of silicon nanocrystals doped with phosphorus and Auger recombination rates are presented. It is shown that the presence of several phosphorus atoms leads to stronger decrease of the optical gap and to growth of the Auger rates comparing to the case of the one atom.

Keywords: silicon, nanocrystal, density functional theory, Auger-process.

Derbeneva Natalia Victorovna — postgraduate student, kurova_natasha@mail.ru

ПРЯМОЕ ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ ДВУХФАЗНОГО ТУРБУЛЕНТНОГО ПОГРАНИЧНОГО СЛОЯ НАД ВЗВОЛНОВАННОЙ ВОДНОЙ ПОВЕРХНОСТЬЮ

О. А. Дружинин, Ю. И. Троицкая

Институт прикладной физики Российской академии наук (ИПФ РАН)

Один из важнейших факторов, влияющих на процессы обмена импульсом в турбулентном воздушном погранслое над водной поверхностью, — наличие мелкодисперсных капель. Согласно данным натурных наблюдений, массовая доля морского аэрозоля растёт с увеличением скорости ветра, прежде всего, за счёт инжекции капель в воздух при микрообрушениях в окрестности гребней волн, и может достигать десятков процентов. Поэтому учёт двухфазности воздушного потока над волнами может существенно влиять на параметризацию процессов турбулентного обмена в прогнозных моделях климата. В данной работе с помощью прямого численного моделирования исследуются свойства воздушного турбулентного погранслоя над волнами, несущего мелкодисперсные капли.

Работа выполнена при поддержке Федеральной целевой программы «Научные и педагогические кадры для инновационной России», контракт № 14.578.21.0033 (2014–2016); проекта академии наук Финляндии «Взаимодействие атмосферы и гидросферы в бассейне Балтийского моря и арктических морей» № 280700 (2014–2017); Российского фонда фундаментальных исследований (проекты № 14-05-00367, 13-05-00865, 14-05-91767, 15-35-20953). Разработка численного метода и численные расчёты, обработка их результатов выполнены при поддержке Российского научного фонда (проекты № 14-17-00667 и 15-17-20009).

Ключевые слова: прямое численное моделирование, турбулентный ветер, двумерные волны на воде, мелкодисперсные капли.

ВВЕДЕНИЕ

Процессы обмена массой, импульсом и теплом между атмосферой и океаном — один из ключевых факторов, влияющих на формирование климата. Моделирование этих процессов основано на процедуре параметризации, связывающей турбулентные потоки с градиентами средних полей плотности и скорости воздуха посредством так называемых балк-формул. Коэффициенты обмена, содержащиеся в балк-формулах, зависят от деталей процесса взаимодействия ветра с волнами, которые и по сей день остаются до конца не выясненными, несмотря на большой объем имеющихся лабораторных и натурных данных [Fairall et al., 2003].

Большой интерес представляют мелкодисперсные капли, взвешенные в воздушном потоке вблизи морской поверхности. Предполагается, что их влияние на несущий воздушный поток при определённых условиях может приводить к существенному снижению сопротивления водной поверхности и, соответственно, к увеличению скорости потока [Kudryavtsev 2006; Kudryavtsev, Makin 2011].

Дружинин Олег Александрович — ведущий научный сотрудник, доктор физико-математических наук, druzhinin@hydro.appl.sci-nnov.ru

Троицкая Юлия Игоревна — заведующая отделом, доктор физико-математических наук, yuliya@hydro.appl.sci-nnov.ru

Данные натуральных и лабораторных экспериментов показывают, что доминирующая роль в инжекции капель в воздушный поток вблизи водной поверхности принадлежит струям, формирующимся при достаточно большой скорости ветра при микрообрушениях (в виде «барашков») в окрестности гребней волн [Andreas et al. 2010; Fairall et al., 2009]. При этом оказывается, что относительно крупные капли (с диаметром около 100 мк) обуславливают основной вклад в массовую долю дисперсной фазы в воздушном потоке. Измерения показывают также, что массовая доля капель растёт с увеличением скорости ветра и может достигать десятков процентов вблизи водной поверхности. Необходимо отметить, однако, что известные экспериментальные данные пока недостаточны для однозначного ответа на вопрос о том, как капли изменяют свойства воздушного потока над взволнованной водной поверхностью, и в частности, коэффициент сопротивления водной поверхности.

В данной работе для исследования влияния мелкодисперсных капель на воздушный поток над волнами используется метод прямого численного моделирования (Direct Numerical Simulation, DNS). До настоящего времени DNS проводилось лишь для случая однофазного воздушного потока [Druzhinin et al., 2012; Sullivan et al., 2000; Yang, Shen, 2010].

В данной статье рассматривается нейтрально-стратифицированный воздушный поток с балковым числом Рейнольдса $Re = 15\,000$ и кругизной волны $ka = 0,2$. Используется эйлерово описание для несущего воздушного потока и лагранжево описание для дисперсной фазы. Считается, что объёмная концентрация капель достаточно мала, и эффектами их коагуляции можно пренебречь. Однако массовая доля капель при этом оказывается значительной, такой что их эффект на несущий воздушный поток конечен (заметен).

1. ФОРМУЛИРОВКА ЗАДАЧИ

Проводится прямое численное моделирование турбулентного воздушного течения Куэтта, несущего дискретные капли, над взволнованной водной поверхностью. Схема численного эксперимента аналогична схеме, рассматриваемой в работе [Druzhinin et al., 2012], где проводилось DNS турбулентного однофазного воздушного потока (рис. 1). Рассматривается декартова система координат с осью x , ориентированной в направлении среднего воздушного потока, осью z , направленной вертикально вверх (против ускорения свободного падения g) и осью y в поперечном направлении, параллельном фронту волны. Анализируется двухмерная поверхностная волна с амплитудой a , длиной λ и фазовой скоростью c , периодическая в направлении x . В данной работе рассматривается кругизна волны $ka = 2\pi a/\lambda = 0,2$.

Метод DNS проводится в системе отсчёта, движущейся с фазовой скоростью волны, т. е. горизонтальная координата x в этой системе отсчёта связана с координатой в лабораторной системе отсчёта x' соотношением $x = x' - ct$. Таким образом, в движущейся системе отсчёта нижняя граница области счета, совпадающая с волной, стационарна. На этой границе ставится условие прилипания, т. е. x - и z -компоненты скорости воздушного потока здесь совпадают с орбитальными скоростями частиц жидкости в волне и не зависят от времени. Рассматривается область счёта с размерами $L_x = 6\lambda$, $L_y = 4\lambda$ и $L_z = \lambda$ в x -, y - и z -направлениях. Течение предполагается периодическим в направлениях x и y . Ставится условие прилипания на верхней (плоской) границе, которая движется с балковой скоростью U_0 в направлении x .

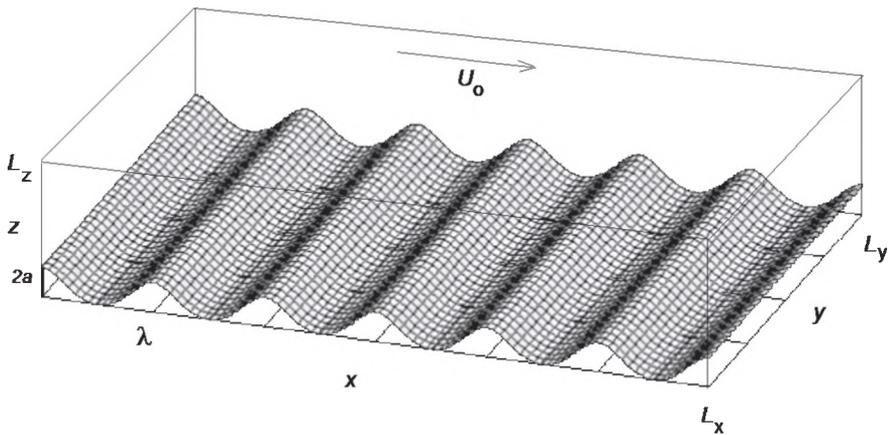


Рис. 1. Схема численного эксперимента

Благодаря этому условию создаётся внешний источник потока импульса за счёт действия вязкого напряжения сдвига, который компенсирует вязкую диссипацию кинетической энергии турбулентности и обеспечивает стационарность течения.

Численное моделирование основано на интегрировании полных трёхмерных уравнений Навье-Стокса для несжимаемой жидкости в безразмерных переменных в виде [Монин, Яглом, 1992]:

$$\frac{\partial U_i}{\partial t} + \frac{\partial(U_i U_j)}{\partial x_j} = -\frac{\partial P}{\partial x_j} + \frac{1}{\text{Re}} \frac{\partial^2 U_i}{\partial x_j \partial x_j} + \sum_{m=1}^{N_d} f_i^m, \quad (1)$$

$$\frac{\partial U_j}{\partial x_j} = 0, \quad (2)$$

где $U_i (i = x, y, z)$ — компоненты скорости; P — давление; f_i^m — сила, с которой m -я капля действует на окружающий её воздух; N_d — общее (постоянное в DNS) число капель. Переменные в выражениях (1), (2) нормированы на длину волны λ , балковую скорость U_0 , давление нормировано на ρU_0^2 (где ρ — плотность воздуха). Балковое число Рейнольдса равно

$$\text{Re} = \frac{U_0 \lambda}{\nu}. \quad (3)$$

Каждая капля рассматривается как твёрдая сфера с плотностью, в δ раз превышающей плотность воздуха (для воды), и диаметром d , много меньшим вязкого пристеночного масштаба ν/u_* , где u_* — скорость трения; ν — кинематическая вязкость воздуха. Это позволяет рассматривать каждую каплю в уравнении (1) как мгновенный точечный источник и аппроксимировать силу, с которой она воздействует на окружающий воздух в данном узле сетки, следующим выражением:

$$f_i^m = -w(r_g, r_m) F_i^m \delta, \quad (4)$$

где F_i^m — сила, действующая на каплю со стороны окружающего воздуха; $w(r, r_m)$ — весовой коэффициент, зависящий от координат капли $r_m = (x_m, y_m, z_m)$ и рассматриваемого узла сетки $r_g = (x_g, y_g, z_g)$ [Druzhinin, 2001]. В рассматриваемом

мом случае (капли воды в воздухе) $\delta = 1000$, и сила F_i^m хорошо описывается Стоксовым законом сопротивления, включающим поправку, связанную с конечностью числа Рейнольдса капли Re_m , в виде [Maxey, Riley, 1983]:

$$F_i^m = 3\pi d\nu\rho\left(U_i(r_m) - V_i^m\right)\left(1 + Re_m^{0,687}\right), \quad (5)$$

где

$$Re_m = \frac{|V^m - U|d}{\nu}. \quad (6)$$

Движение каждой капли отслеживается в лагранжевых координатах согласно уравнениям движения:

$$\frac{dr_i^m}{dt} = V_i^m(t), \quad (7)$$

$$\frac{dV_i^m}{dt} = \frac{1}{\Omega_d} F_i^m(t) - \frac{1}{Fr_d^2} \delta_{ij}, \quad (8)$$

где $\Omega_d = \pi d^3/6$ — объём капли; δ_{ij} — символ Кронекера; число Фруда капли равно

$$Fr_d = \frac{(g\lambda)^{1/2}}{U_0}. \quad (9)$$

Для определения значения скорости воздуха в точке пространства, занимаемой каплей, использовалась эрмитова интерполяция 4-го порядка точности.

Для того чтобы избежать эффектов, связанных с сильной геометрической нелинейностью при интегрировании (1)–(3) из-за волнообразной нижней границы (см. рис. 1), используется конформное преобразование координат, переводящее плоскость (x, z) в декартовых координатах в плоскость (ξ, η) в криволинейных координатах, а именно:

$$x = \xi - a \exp(-k\eta) \sin k\xi, \quad (10)$$

$$z = \eta + a \exp(-k\eta) \cos k\xi. \quad (11)$$

При этом преобразовании волновая граница $z_b(x) = a \cos k\xi(x)$ переходит в плоскую границу $\eta = 0$. В дополнение используется также отображение по вертикальной координате η , обеспечивающее достаточное разрешение больших градиентов полей скорости и температуры благодаря сгущению узлов сетки вблизи нижней и верхней границ [Druzhinin et al., 2012].

Уравнения (1)–(3) решались в прямоугольной области с размерами $0 < \xi < 6$, $0 < \eta < 4$, и $0 < \eta < 1$ при помощи метода конечных разностей второго порядка на разнесённой сетке, состоящей из $360 \times 240 \times 180$ узлов. Интегрирование по времени осуществлялось с использованием метода Адамса-Башфорфа второго порядка точности в две стадии на каждом временном шаге n . На нижней и верхней границах накладывались условия прилипания (Дирихле) для скорости воздушного потока. На боковых границах, при $x = 0,6$ и $y = 0,4$, на все поля накладывались периодические граничные условия.

При совпадении вертикальной координаты капли с верхней или нижней границами производилась её новая инжекция в поток. При задании исходных координат и скорости капли принимались во внимание механизмы генерации капель под действием ветра по данным лабораторных экспериментов [Andreas et al.,

2010]. Таким образом, капли инжектировались в окрестности наветренных гребней волны в начале буферного подслоя (около $10...20z_*$) равномерно вдоль фронта волны со скоростью, равной скорости воздуха.

Течение инициировалось как слабо-возмущённое ламинарное течение Куэтта, $U_i = z\delta_{ix} + u_{if}$ ($i = x, y, z$), где u_{if} — бездивергентное, изотропное, однородное случайное поле с широким спектром и амплитудой порядка 0,05. Интегрирование во времени сначала проводилось до окончания переходных процессов, связанных с выбором начального условия для воздуха. Далее, в момент времени $t = 100$, в поток инжектировались капли со случайным равномерным пространственным распределением и со скоростями, равными локальным скоростям частиц воздуха.

По достижении статистически стационарного режима течения проводилось осреднение по времени и координате y . Для усреднения по времени брались поле скорости и координаты капель в дискретные моменты $t_k, k = 1, \dots, 1000$ с интервалом $t_{k+1} - t_k = 0,2$. Помимо осреднения по y и t рассчитывались также фазовые (оконные) средние по длине волны.

Таким образом, для данного поля $f(x, y, z, t)$ определялись фазовое среднее и дисперсия:

$$\langle f \rangle(x, z) = \frac{1}{6N_t N_y} \sum_{j=1}^{N_y} \sum_{k=1}^{N_t} \sum_{n=0}^5 f(x + n\lambda, y_j, z, t_k), \quad (12)$$

$$\langle f^2 \rangle(x, z) = \frac{1}{6N_t N_y} \sum_{j=1}^{N_y} \sum_{k=1}^{N_t} \sum_{n=0}^5 f^2(x + n\lambda, y_j, z, t_k), \quad (13)$$

где $N_y = 240, N_t = 500$ и $0 < x < 1$. Далее флуктуация поля f определялась как:

$$f'(x, z) = \left| \langle f^2 \rangle(x, z) - \langle f \rangle^2(x, z) \right|^{1/2} \quad (14)$$

(в случае расчёта потока, например $\langle U_x U_z \rangle$, в выражение (10) подставляется произведение $\langle U_x U_z \rangle$ вместо f^2 , и в (11) — разность $\left| \langle U_x U_z \rangle - \langle U_x \rangle \langle U_z \rangle \right|$).

Средний вертикальный профиль поля $f, [f](z)$, рассчитывался осреднением фазового среднего по продольной координате:

$$[f](z) = \frac{1}{N_x} \sum_{i=1}^{N_x/6} \langle f \rangle(x_i, z). \quad (15)$$

Вертикальный турбулентный поток импульса $\tau_m(z)$ определяется в виде

$$\tau(z) = \left| \langle U_x U_z \rangle - \langle U_x \rangle \langle U_z \rangle \right|. \quad (16)$$

Таким образом, динамическая скорость (трения), u_* , определяется из асимптотического значения потока импульса на достаточно большом расстоянии от поверхности в виде

$$u_* = \sqrt{-\tau}. \quad (17)$$

Численный алгоритм для воды верифицировался путём сравнения результатов DNS для турбулентного погранслоя над плоской поверхностью при различных балковых числах Ричардсона и числах Рейнольдса. Результаты DNS сравнивались с известными численными и экспериментальными данными [Druzhinin et al., 2012].

2. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

Численное моделирование проводилось для крутизны волны $ka = 0,2$, отношения фазовой скорости к динамической скорости (возраста волны) в интервале $c/u_* = 0,05$ (типичного для сильных, т.е. эффективно генерирующих капли, ветров в натуральных условиях) и балковом числе Рейнольдса $Re = 15\,000$. Диаметр капли задавался достаточно малым (равным $d = 0,00077$ в безразмерных единицах) по сравнению с вязким масштабом (около $0,002$). При этом число Фруда задавалось равным $Fg_d = 0,02$ таким, что отношение скорости гравитационного оседания капли в покоящемся воздухе

$\left[V_g = g \frac{d^2}{18\nu} \delta \right]$ к скорости трения (помноженной на постоянную Кармана, $\kappa = 0,4$) оказывалось порядка единицы, $V_g / \kappa u_* \approx 1$.

Этот режим в типичных лабораторных и натуральных экспериментах соответствует каплям с диаметром порядка $100\ \mu\text{м}$. Общее максимальное число капель достигало $16\ \text{млн}$, что при данном диаметре капли соответствует исходной объёмной концентрации порядка $1,6 \cdot 10^{-4}$ и массовой доли порядка $16\ \%$.

Мгновенные распределения модуля завихренности и расположения капель в центральных плоскостях (x, z) , (y, z) и (x, y) при $z = 0,04$, полученные в DNS в момент времени $t = 300$, представлены на рис. 2. Видны типичные для турбулентного погранслоя вихревые структуры (так называемые Λ -вихри), формирующиеся в окрестности гребней волн и «уходящие» от водной поверхности вглубь воздушного потока. Капли переносятся этими структурами и, благодаря своей инерции, кластеризуются на их периферии — эффект, хорошо известный в литературе под названием *preferential accumulation*.

На рис. 3 дана траектория отдельно взятой капли, неоднократно падающей на поверхность воды и вновь инжектируемой в поток, полученная в DNS в течение рассматриваемого временного интервала ($100 < t < 300$). Видно, что капля преимущественно инжектируется с наветренного склона гребня волны вблизи вершины и падает на поверхность в окрестности подошвы. Происходят также выбросы капли вглубь потока, где влияние гравитационного оседания на её траекторию относительно мало.

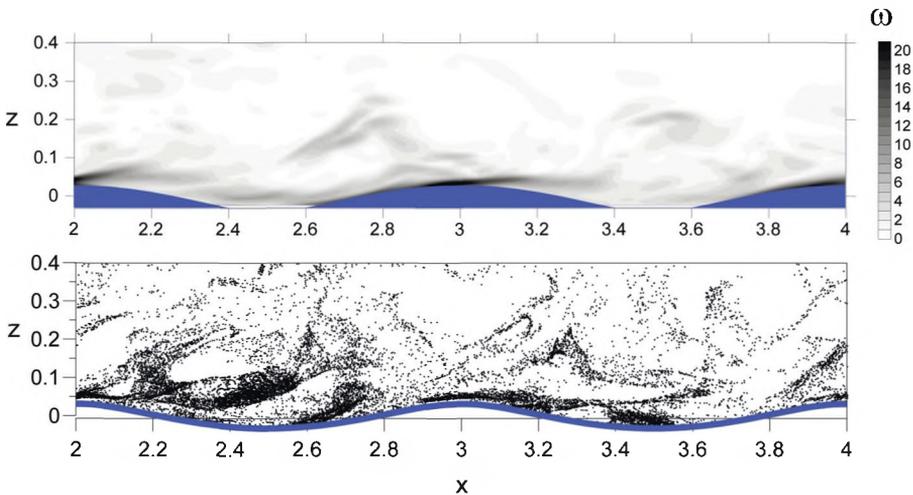


Рис. 2а. Распределение модуля завихренности (верхняя панель) и положение капель в (x, z) -плоскости при $y = 0$ в DNS в момент времени $t = 300$

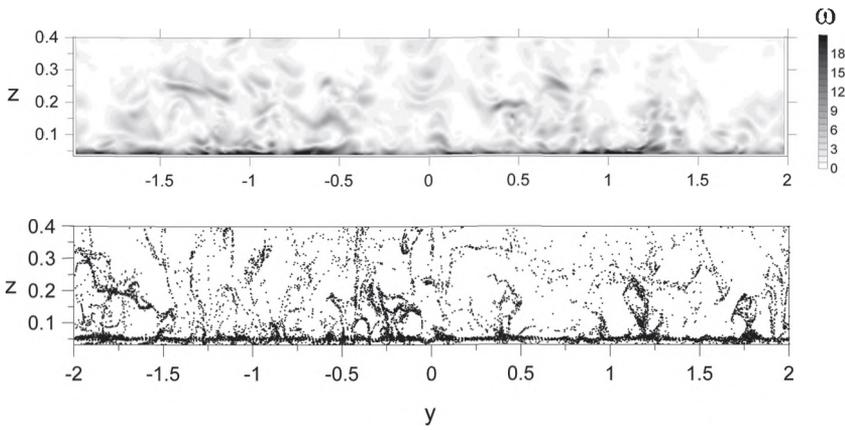


Рис. 26. Распределение модуля завихренности (верхняя панель) и положение капель в (y, z) -плоскости при $x = 3$ в DNS в момент времени $t = 300$

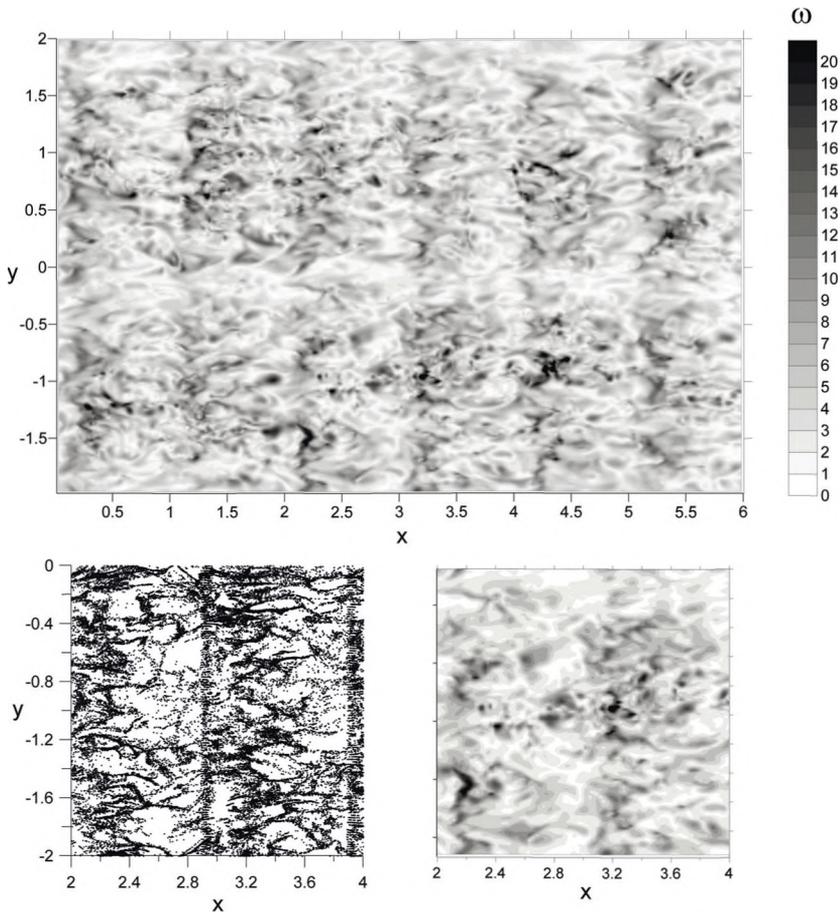


Рис. 2а. Распределение модуля завихренности (верхняя и нижняя правая панели) и положение капель (нижняя левая панель) в (x, y) -плоскости при $z = 0,04$ в DNS в момент времени $t = 300$

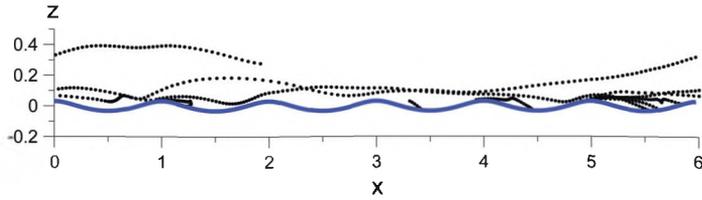


Рис. 3. Трассировка капли с начальными координатами $x = 3,31$, $y = 0,335$ и $z = 0,0337$, полученная в DNS на временном интервале $100 < t < 300$

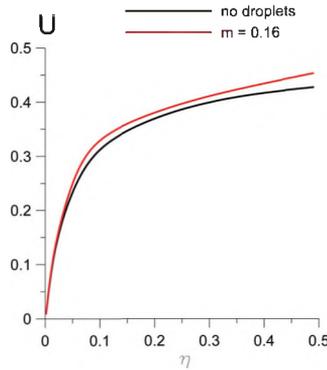


Рис. 4. Вертикальный профиль средней скорости однофазного и двухфазного потоков. Массовая концентрация капель 0,16

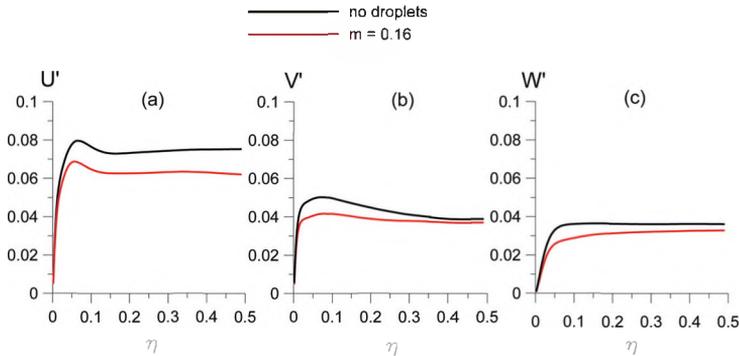


Рис. 5. Профили флуктуаций (дисперсий) x -, y - и z -компонент скорости в однофазном и двухфазном потоках

На рис. 4 представлены вертикальные профили средней скорости воздуха, $U(\eta)$, полученные в DNS для случая однофазного (без капель) и двухфазного потоков. Видно, что под действием капель происходит ускорение воздушного потока в рассматриваемой области пограничного слоя.

Профили флуктуаций x -, y - и z -компонент скорости воздуха, полученные для однофазного и двухфазного воздушных потоков, показаны на рис. 5. Видно, что под действием капель амплитуды флуктуаций всех компонент скорости уменьшаются. На рис. 6 также видно, что в двухфазном потоке уменьшается и турбулентный поток импульса, τ , рассчитанный согласно (16).

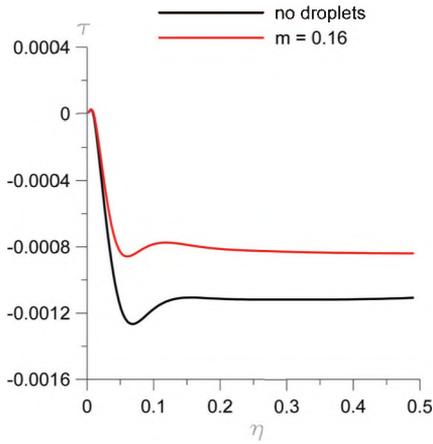


Рис. 6. Профили турбулентного потока импульса в однофазном и двухфазном потоках

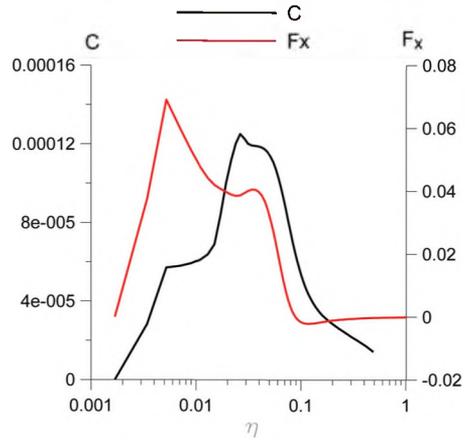


Рис. 7. Профили концентрации капель и x-компоненты силы, с которой капли действуют на воздушный поток

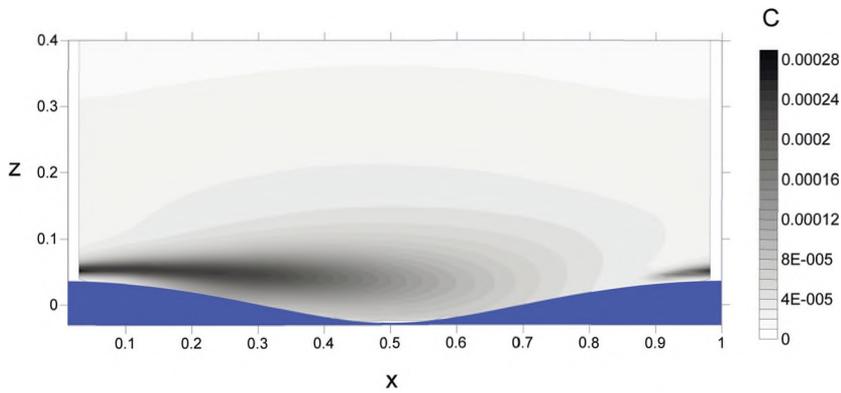


Рис. 8. Фазовое среднее поле концентрации капель

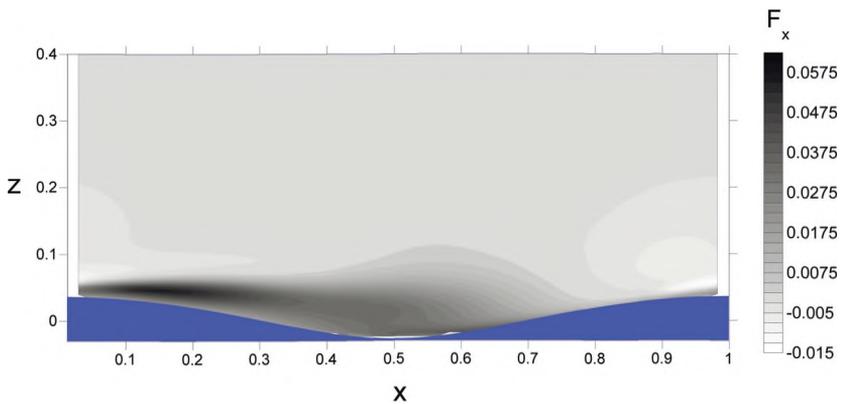


Рис. 9. Фазовое среднее поле x-компоненты силы, с которой капли воздействуют на несущий воздушный поток

В DNS рассчитывались поля объёмной концентрации капель C . Концентрация капель в рассматриваемом узле сетки $r_g = (x_g, y_g, z_g)$, $C(r_g)$, рассчитывалась как

$$C(r_g) = \frac{\pi d^3}{6\Delta x^2 \Delta z} \sum_{m=1}^{N_d} w(r_g, r_m), \quad (18)$$

где суммирование ведётся по всем каплям, $w(r_g, r_m)$ — те же весовые коэффициенты, что и в выражении (4); $r_m = (x_m, y_m, z_m)$ — координаты капли. На рис. 7 показаны средние вертикальные профили концентрации $C(z)$ и x -компоненты силы, с которой капли действуют на воздушный поток, рассчитанной согласно (1), (4), (5).

Видно, что капли кластеризуются вблизи водной поверхности при $0,01 < z < 0,1$. Хорошо выраженный максимум концентрации расположен на расстоянии порядка амплитуды волны, в области буферного слоя ($z \approx 0,03$, $z^+ \approx 15$). Максимумы силы воздействия капель на поток приходится на область вязкого подслоя ($z \approx 0,007$, $z^+ \approx 4$) и буферного слоя.

Оконные (фазовые) средние поля концентрации и x -компоненты силы, с которой капли действуют на воздушный поток, рассчитанные согласно (12), представлены на рис. 8 и 9.

На рис. 8 и 9 видно, что области максимумов концентрации капель и силы, с которой они воздействуют на поток, хорошо пространственно скоррелированы и сосредоточены на подветренном склоне гребней поверхностной волны. Здесь капли ускоряют поток, так как они опережают, благодаря своей инерции, замедляющиеся в рассматриваемой области частицы воздуха. Напротив, в области наветренного склона вблизи гребня капли замедляют воздушный поток, так как в этой области частицы воздуха ускоряются.

Ещё одной важной характеристикой двухфазного потока является средняя скорость гравитационного оседания капель. Для определения эйлерова поля скорости капель сначала рассчитывалась мгновенная скорость в каждом узле сетки $r_g = (x_g, y_g, z_g)$ в виде

$$V_z(r_g) = \frac{\pi d^3}{6\Delta x^2 \Delta z} \sum_{m=1}^{N_d} w(r_g, r_m) V_z^m, \quad (19)$$

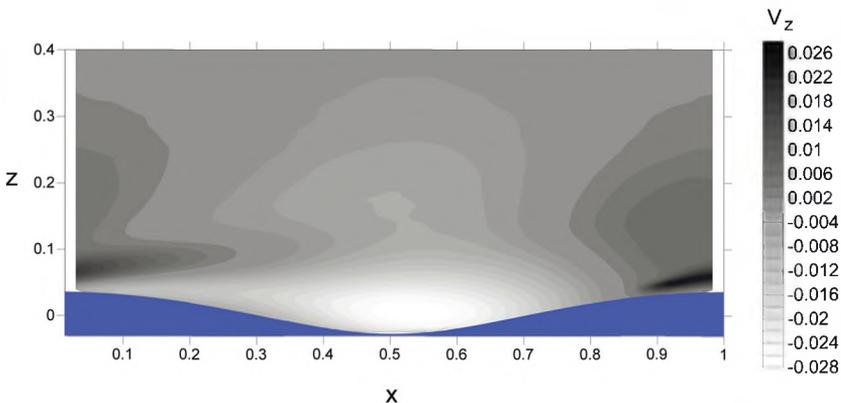


Рис. 10. Фазовое среднее поле z -компоненты скорости капель

где суммирование ведётся по всем каплям, $w(r_g, r_m)$ — те же весовые коэффициенты, что и в выражении (4); $r_m = (x_m, y_m, z_m)$ — координаты капли. Далее согласно (12) определялось фазовое среднее поле вертикальной компоненты скорости капель. Полученное поле средней вертикальной скорости капель представлено на рис. 10. Видно, что капли оседают в основном вблизи подошв волн и уходят вглубь потока в окрестности гребней волн.

ЗАКЛЮЧЕНИЕ

Впервые выполнено прямое численное моделирование (DNS) двухфазного воздушного потока над взволнованной водной поверхностью при массовой концентрации мелкодисперсных капель 16 %. Параметры дисперсной фазы (капель) в DNS задавались исходя из известных экспериментальных данных. Результаты показывают, что при инъекции капель в поток в окрестности наветренного склона поверхностной волны вблизи гребней основным эффектом воздействия капель на поток является увеличение скорости ветра. При этом под действием капель уменьшаются турбулентные флуктуации и турбулентный поток импульса по сравнению с однофазным потоком. Несмотря на существенный вклад эффектов, связанных с инерцией капель, результаты DNS в целом согласуются с результатами известных феноменологических моделей, в которых инерция капель не учитывается, а принимается во внимание лишь их гравитационное оседание [Kudryavtsev, Makin, 2011].

ЛИТЕРАТУРА

- [Монин, Яглом, 1992] Монин А. С., Яглом А. М. Статистическая гидромеханика. Т. 1. СПб.: Гидрометеиздат, 1992. 694 с.
- [Andreas et al., 2010] Andreas E. L., Jones K. F., Fairall C. W. Production velocity of sea spray droplets // J. Geophys. Res. 2010. V. 115. C12065. doi:10.1029/2010JC006458.
- [Druzhinin, 2001] Druzhinin O. A. The influence of particle inertia on the two-way coupling and modification of isotropic turbulence by microparticles // Phys. Fluids. 2001. V. 13. P. 3738–3755.
- [Druzhinin et al., 2012] Druzhinin O. A., Troitskaya Yu. I., Zilitinkevich S. S. Direct numerical simulation of a turbulent wind over a wavy water surface // J. Geophys. Res. 2012. V. 117. C00J05. doi: 10.1029/2011JC007789.
- [Fairall et al., 2003] Fairall C. W., Bradley E. F., Hare J. E. Bulk parameterization of air-sea fluxes: updates and verification for the COARE algorithm // J. Climate. 2003. V. 16. P. 571–591.
- [Fairall et al., 2009] Fairall C. W., Banner M. L., Pierson W. L., Asher W., Morrison R. P. Investigation of the physical scaling of sea spray spume droplet production // J. Geophys. Res. 2009. V. 114. C10001. doi:10.1029/2008JC004918.
- [Kudryavtsev, 2006] Kudryavtsev V. N. On the effect of sea drops on the atmospheric boundary layer // J. Geophys. Res. 2006. V. 111. C07020. doi:10.1029/2005JC002970.
- [Kudryavtsev, Makin, 2011] Kudryavtsev V. N., Makin V. K. Impact of Ocean Spray on the Dynamics of the Marine Atmospheric Boundary Layer // Boundary-Layer Meteorol. 2011. V. 140. P. 383–410. doi: 10.1007/s10546-011-9624-2.
- [Maxey, Riley, 1983] Maxey M. R., Riley J. J. Equation of motion for a small rigid sphere in a non-uniform flow // Phys. Fluids. 1983. V. 26. P. 49–60.

- [Richter, Sullivan, 2013] *Richter D. H., Sullivan P. P.* Sea surface drag and the role of spray // *Geophys. Res. Lett.* 2013. V. 40. No. 1–5. doi:10.1002/grl.50163.
- [Sullivan et al., 2000] *Sullivan P. P., McWilliams J. C., Moeng C.-H.* Simulation of turbulent flow over idealized water waves // *J. Fluid Mech.* 2000. V. 404. P. 47–85.
- [Yang, Shen, 2010] *Yang D., Shen L.* Direct-simulation-based study of turbulent flow over various waving boundaries // *J. Fluid Mech.* 2010. V. 650. P. 131–180.

DIRECT NUMERICAL SIMULATION STUDY OF A TURBULENT STABLY STRATIFIED AIR FLOW ABOVE THE WAVY WATER SURFACE

O. A. Druzhinin, Yu.I. Troitskaya

Institute of Applied Physics Russian Academy of Sciences (IAP RAS)

The presence of droplets in the air flow above the waved water surface can significantly affect the turbulent momentum exchange between the wind and surface waves. Field experiments show that mass fraction of sea aerosol in the vicinity of the water surface increases with the wind speed, mainly due to micro-breaking events in the vicinity of the wave crests, and can reach up to dozens percent. Therefore taking into account the droplet phase in the air flow above waves can considerably modify parameterization of turbulent exchange processes in prognostic models of climate. In the present work, we use Direct Numerical Simulation to study the properties of turbulent air-flow above the waved water surface taking into account airborne droplets.

Keywords: direct numerical simulation, turbulent wind, two-dimensional water waves, droplets.

Druzhinin Oleg Aleksandrovich — leading scientist, doctor of physical and mathematical sciences, druzhinin@hydro.appl.sci-nnov.ru

Troitskaya Yuliya Igorevna — head of department, doctor of physical and mathematical sciences, yuliya@hydro.appl.sci-nnov.ru

ДИНАМИЧЕСКИЕ СВОЙСТВА ОДНОГО КЛАССА ИМПУЛЬСНЫХ СИСТЕМ

Л. И. Ивановский

Ярославский государственный университет им. П. Г. Демидова (ЯрГУ им. П. Г. Демидова)

В работе рассматривается математическая модель динамической системы, представляющая собой цепочку из трёх диффузионно связанных в кольцо сингулярно возмущённых нелинейных дифференциальных уравнений с запаздыванием. Применительно к ним изучаются вопросы существования и устойчивости релаксационных периодических решений на основании бифуркационного анализа специального двумерного отображения. Особое внимание уделяется числу сосуществующих устойчивых режимов.

Исследование выполнено при поддержке Российского научного фонда (проект № 14-21-00158).

Ключевые слова: бифуркации, релаксационные автоколебания, устойчивые режимы.

ВВЕДЕНИЕ

Рассмотрим цепочку связанных в кольцо сингулярно возмущённых осцилляторов с запаздыванием:

$$\left. \begin{aligned} \dot{u}_j &= d(u_{j+1} - 2u_j + u_{j-1}) + \lambda \left[-1 + \alpha f(u_j(t-1)) - \beta g(u_j) \right] u_j, \\ j &= \overline{1, m}, \quad u_0 = u_m, \quad u_1 = u_{m+1}, \end{aligned} \right\} \quad (1)$$

где $m \geq 2$, $\lambda \gg 1$, $\beta > 0$, $\alpha > 1 + \beta$, а $u_j = u_j(t) > 0$ моделируют мембранные потенциалы нейронов. Гладкие функции $f(u), g(u) \in C^2(\mathbb{R}_+)$, где $\mathbb{R}_+ = \{u \in \mathbb{R} : u \geq 0\}$, удовлетворяют условиям

$$0 < \beta g(u) < \alpha, \quad f(0) = g(0) = 1, \quad \forall u \in \mathbb{R}_+;$$

$$f(u), g(u), u f'(u), u g'(u) = O(1/u) \quad \text{при } u \rightarrow +\infty.$$

В статьях [Глызин и др., 2011а, б, 2012а, б, 2013] было показано, что с помощью замены вида

$$u_1 = \exp\left(\frac{x}{\varepsilon}\right), \quad u_j = \exp\left(\frac{x}{\varepsilon} + \sum_{k=1}^{j-1} y_k\right), \quad j = \overline{2, m}, \quad \varepsilon = \frac{1}{\lambda} \ll 1,$$

где x, y_1, \dots, y_{m-1} — новые переменные, систему (1) можно свести к предельной системе обыкновенных дифференциальных уравнений без малого параметра, но с импульсными воздействиями. Рассмотрим данную задачу в случае трёх сингулярно возмущённых осцилляторов ($m = 3$):

$$\left\{ \begin{aligned} \dot{y}_1 &= d(e^{y_2} + e^{-y_1} - e^{y_1} - e^{y_2+y_2}), \\ \dot{y}_2 &= d(e^{-y_1-y_2} + e^{-y_2} - e^{y_2} - e^{-y_1}). \end{aligned} \right. \quad (2)$$

$$\left. \begin{aligned} y_j(+0) &= \frac{\alpha-1}{\alpha-\beta-1} y_j(-0), y_j(1+0) = y_j(1-0) - \frac{\alpha}{\alpha-1} y_j(+0), \\ y_j(\alpha+0) &= (1+\beta) y_j(\alpha-0), \\ y_j(\alpha+1+0) &= y_j(\alpha+1-0) - \frac{\alpha}{1+\beta} y_j(\alpha+0), \quad j=1,2. \end{aligned} \right\} \quad (3)$$

Отметим, что функции $y_1(t)$ и $y_2(t)$ характеризуют фазовые сдвиги между компонентами системы (1).

Введём в рассмотрение отображение следующего вида:

$$\tilde{O}(z): \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \rightarrow \begin{pmatrix} y_1(T_0) \\ y_2(T_0) \end{pmatrix}, \quad (4)$$

где $y_1(-0) = z_1$, $y_2(-0) = z_2$, а величина T_0 определяет главную часть периода устойчивого цикла одиночного осциллятора системы (1). Оно сопоставляет начальным условиям системы (2) решение с координатами y_j в момент времени $t = T_0$.

В статьях [Глызин и др., 2011а, б, 2012а, б, 2013] было доказано, что экспоненциально устойчивым неподвижным точкам отображения (4) соответствуют орбитально асимптотически устойчивые циклы системы (1). Другими словами, для того, чтобы говорить об устойчивых циклах, достаточно изучить неподвижные точки модельного отображения. Они и являлись объектами исследования.

Асимптотический анализ отображения (4) показывает, что при достаточно малых значениях параметра d оно имеет как минимум три устойчивые неподвижные точки, при этом нулевое состояние равновесия устойчиво для любых значений d . Задача исследования состоит в определении таких значений параметров α и β , при которых отображение (4) имеет наибольшее число устойчивых неподвижных точек. Поскольку сделать это в полной мере с использованием лишь аналитического аппарата затруднительно, поиск устойчивых состояний осуществлялся с помощью специально разработанной программы. Вычисление координат неподвижных точек отображения (4) в этом приложении осуществлялось параллельно, на независимых потоках центрального процессора.

ПОИСК НЕПОДВИЖНОЙ ТОЧКИ МОДЕЛЬНОГО ОТОБРАЖЕНИЯ

Для получения формул решения задачи (2) на промежутках между импульсами удобно рассмотреть систему линейных дифференциальных уравнений

$$\dot{u} = d \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix} u, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix},$$

общее решение которой имеет следующий вид:

$$u = \alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 e^{-3t} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + e^{-3t} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Полученное представление и решение системы (2) связаны простыми формулами $y_1 = \ln(u_2/u_1)$, $y_2 = \ln(u_3/u_2)$, из которых имеем

$$y_1 = \ln \left[\frac{\alpha_1 - \alpha_2 e^{-3t}}{\alpha_1 + \alpha_2 e^{-3t} + e^{-3t}} \right], \quad y_2 = \ln \left[\frac{\alpha_1 - e^{-3t}}{\alpha_1 - \alpha_2 e^{-3t}} \right].$$

Если изучается начальная задача Коши (2) с начальными условиями (3), нетрудно найти коэффициенты α_1, α_2 :

$$\alpha_1 = \frac{e^{-3t_0} (e^{z_{01} + z_{02}} + e^{z_{01}} + 1)}{e^{z_{01}} - 2e^{z_{01} + z_{02}} + 1}, \quad \alpha_2 = \alpha_1 e^{3t_0} (1 - e^{-z_{02}}) + e^{-z_{02}}.$$

Выбирая в качестве величин z_{01}, z_{02} координаты точек, полученных после очередного импульса в моменты времени $0, 1, \alpha, \alpha + 1$ нетрудно получить решение импульсной системы (2), (3) на промежутке длины периода.

Если координаты начальной и конечной точек отображения (4) оказывались достаточно близкими друг к другу, предполагалось, что в этом случае найдено устойчивое состояние равновесия. В противном же случае, вектор $(y_1(T_0), y_2(T_0))^T$ считали начальным и для него выполняли описанную выше последовательность операций.

РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ИССЛЕДОВАНИЯ

При относительно малых значениях d не трудно обнаружить состояния равновесия, предсказанные аналитически. Для обнаружения других неподвижных точек величины α и β фиксировались, а значение d менялось. В результате исследования были получены значения начальных параметров, для которых обнаруживалось одновременное существование семи устойчивых состояний равновесия.

Далее, как и в статьях [Ивановский, Самсонов, 2014, 2015], разберём два основных бифуркационных сценария.

Случай $\alpha = 1,9, \beta = 0,1$. При изменении параметра d для заданных значений α и β численный анализ отображения (3) позволяет получить следующий сценарий фазовых перестроек:

- при $d < d_1; d_1 \approx 0,003$ модельное отображение имеет семь устойчивых неподвижных точек и двенадцать неустойчивых. Схематическое изображение фазового портрета отображения Φ для этого случая можно увидеть на рис. 1а. Буквами S_j обозначены устойчивые, а U_j — неустойчивые неподвижные точки;
- при $d = d_1$ три пары симметричных друг относительно друга седла U_7 и U_{12} , U_8 и U_9 , U_{10} и U_{11} одновременно сливаются с устойчивыми узлами S_1 , S_3 и S_5 , отбирая у них устойчивость, и тем самым, образуя седла U'_1 , U'_3 и U'_5 соответственно;
- при $d_1 < d < d_2; d_2 \approx 0,006$ отображение имеет четыре устойчивые неподвижные точки и девять неустойчивых. Фазовый портрет в данном случае имеет вид, изображенный на рис. 1б.
- при $d = d_2$ неустойчивые узлы U'_1 , U'_3 и U'_5 одновременно сливаются с неустойчивыми седлами U_1 , U_3 и U_5 соответственно и пропадают;
- при $d_2 < d < d_3; d_3 \approx 0,021$ отображение имеет четыре устойчивые неподвижные точки и три неустойчивые, как это представлено на рис. 2;
- последняя бифуркация происходит при $d = d_3$. Устойчивые узлы S_2 , S_4 и S_6 , сливаясь с седлами U_2 , U_4 и U_6 соответственно, пропадают. Тем самым, при $d > d_4$ отображение имеет лишь одно единственное нулевое устойчивое состояние.

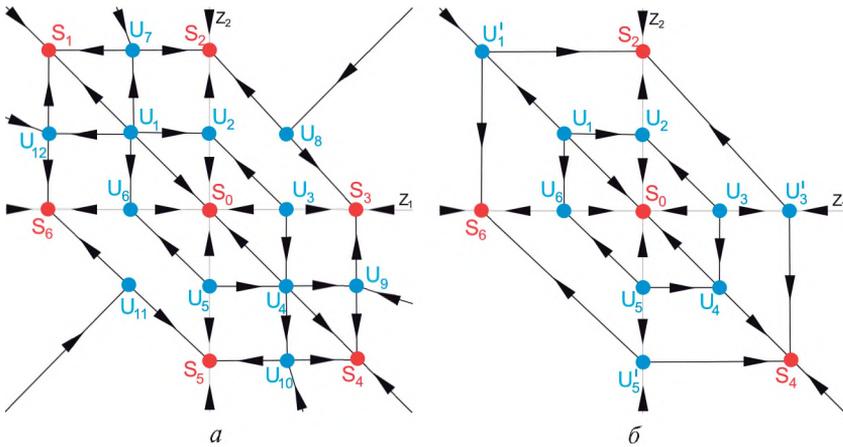


Рис. 1. Фазовые портреты отображения: $a - d < d_1$; $b - d < d_2$

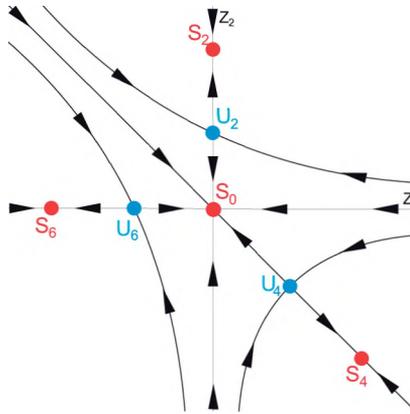


Рис. 2. Фазовый портрет отображения при $d_2 < d < d_3$

Случай $\alpha = 5,0$, $\beta = 0,4$. При изменении параметра d для заданных величин α и β получается следующая последовательность бифуркаций:

- при $d < d_1$; $d_1 \approx 0,021$ отображение (3) имеет семь устойчивых неподвижных точек и двенадцать неустойчивых. Схематическое изображение фазового портрета отображения Φ в данном случае можно увидеть на рис. 3а;
- при $d = d_1$ три пары симметричных друг относительно друга седла U_7 и U_8 , U_9 и U_{10} , U_{11} и U_{12} одновременно сливаются с устойчивыми узлами S_2 , S_4 и S_6 , отбирая у них устойчивость, и, тем самым, образуя седла U'_2 , U'_4 и U'_6 соответственно;
- при $d_1 < d < d_2$; $d_2 \approx 0,031$ отображение имеет четыре устойчивые неподвижные точки и девять неустойчивых. Фазовый портрет в данном случае имеет вид, изображенный на рис. 3б;
- при $d = d_2$ неустойчивые узлы U'_2 , U'_4 и U'_6 одновременно сливаются с неустойчивыми седлами U_2 , U_4 и U_6 соответственно, и пропадают;
- при $d_2 < d < d_3$; $d_3 \approx 0,058$ отображение имеет четыре устойчивые неподвижные точки и три неустойчивые, как это представлено на рис. 4;

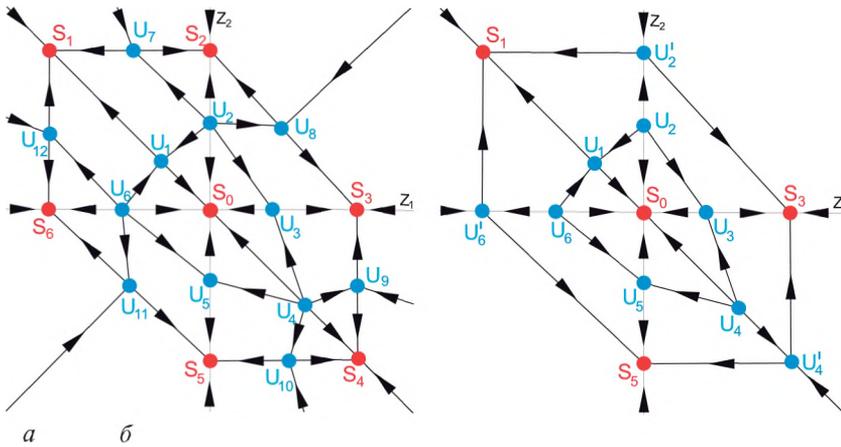


Рис. 3. Фазовые портреты отображения: $d < d_1$; $d_1 < d < d_2$

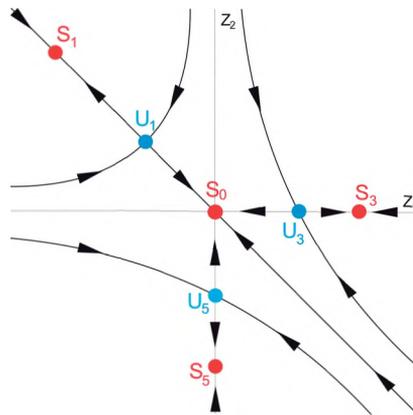


Рис. 4. Фазовый портрет отображения при $d_2 < d < d_3$

- последняя бифуркация происходит при $d = d_3$. Устойчивые узлы S_1 , S_3 и S_5 , сливаясь с седлами U_1 , U_3 и U_5 соответственно, пропадают. Тем самым, при $d > d_4$ отображение имеет лишь одно единственное нулевое устойчивое состояние.

ЗАКЛЮЧЕНИЕ

Благодаря исследованию модельного отображения с помощью разработанного приложения, осуществляющего поиск неподвижных состояний равновесия отображения (4), были получены значения начальных параметров, при которых возможно одновременное сосуществование семи устойчивых режимов. Также были изучены бифуркационные сценарии, происходящие в фазовом пространстве модельного отображения.

ЛИТЕРАТУРА

- [Глызин и др., 2011a] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Релаксационные автоколебания в нейронных системах I // Дифференциальные уравнения. 2011. Т. 47. № 7. С. 919–932.
- [Глызин и др., 2011б] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Релаксационные автоколебания в нейронных системах II // Дифференциальные уравнения. 2011. Т. 47. № 12. С. 1675–1692.
- [Глызин и др., 2012а] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Релаксационные автоколебания в нейронных системах III // Дифференциальные уравнения. 2012. Т. 48. № 2. С. 155–170.
- [Глызин и др., 2012б] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Дискретные автоволны в нейронных системах // Журнал вычислительной математики и математической физики. 2012. Т. 52. № 5. С. 840–858.
- [Глызин и др., 2013] *Глызин С. Д., Колесов А. Ю., Розов Н. Х.* Моделирование эффекта взрыва в нейронных системах // Математические заметки. 2013. Т. 93. № 5. С. 682–699.
- [Ивановский, Самсонов, 2014] *Ивановский Л. И., Самсонов С. О.* Фазовые перестройки одной двумерной динамической системы с импульсным воздействием // Моделирование и анализ информ. систем. 2014. Т. 21. № 6. С. 179–181.
- [Ивановский, Самсонов, 2015] *Ивановский Л. И., Самсонов С. О.* Динамика одного двумерного отображения и устойчивые режимы сингулярно возмущённой системы нейронного типа // Вычислительная техника в естественных науках. Методы суперкомпьютерного моделирования. Ч. 2. 2015. С. 121–132.

DYNAMIC PROPERTIES OF ONE CLASS OF IMPULSE SYSTEMS

L. I. Ivanovsky

P.G. Demidov Yaroslavl State University (P.G. Demidov YarSU)

Let us consider a mathematical model of dynamic system, which is presented as a ring of three diffused connected and singularly perturbed nonlinear differential equations with a delay. In the further text researched the questions of existence and stability of relaxational periodic solutions of this system due to a bifurcational analysis of modeling two-dimensional mapping. Also the special attention is paid to the number of coexisting stable regimes.

Keywords: bifurcations, relaxational self-oscillations, stable regimes.

Ivanovsky Leonid Igorevich — postgraduate student, laboratory researcher at Scientific Center in Chernogolovka Russian Academy of Sciences, leon19unknown@gmail.com

SCARP: МАСШТАБИРУЕМАЯ И ОТКАЗОУСТОЙЧИВАЯ ПЛАТФОРМА ЛЁГКОЙ ВИРТУАЛИЗАЦИИ ДЛЯ ИНТЕНСИВНОЙ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ И ОБЛАЧНЫХ ПРИЛОЖЕНИЙ

Коноплев В. В., Боярский М. Н., Назиров Р. Р.

Институт космических исследований Российской академии наук (ИКИ РАН)

В настоящей работе представлена архитектура и реализация масштабируемого и отказоустойчивого комплекса SCARP, созданного в ИКИ РАН для виртуализации научных приложений, требующих интенсивной коллективной работы с общими наборами данных. Данный комплекс использует свободно распространяемое открытое программное обеспечение и стандартные широко распространённые аппаратные компоненты, что делает проект экономически привлекательным. Использование технологий лёгкой виртуализации позволяет предоставить пользователям одновременно индивидуальное системное окружение и производительность, близкую к производительности физического сервера. В частности, результаты тестов ввода-вывода между виртуальной машиной и коллективным разделяемым хранилищем оказались соизмеримы с локальным вводом-выводом отдельного физического сервера. Комплекс может быть также использован в качестве технологической платформы для развёртывания облачных сервисов для запуска корпоративных и научных приложений.

Ключевые слова: виртуализация, облачные вычисления, OpenVZ, Lustre, ZFS, параллельные файловые системы.

ВВЕДЕНИЕ

Технологии виртуализации вычислений уже давно достигли того уровня развития, при котором этот вопрос переходит из исследовательской области в область прямой экономической целесообразности. Виртуализация позволяет свободно менять физическую вычислительную платформу и легко переносить прикладные задачи в облачные среды. Неслучайно перенос задач в вычислительные облака приобретает все большую популярность в научном сообществе.

Рассматривая возможность масштабного перевода научных вычислительных задач в облачную среду, следует иметь в виду специфику традиционных облачных сервисов. В частности, в экономическую модель предоставления облачных услуг часто заложено предположение о частичной загруженности пользовательских виртуальных машин. В то же время консолидация значительного количества вычислительных ресурсов, а также динамическое распределение между ними пользовательских задач позволяет поставщикам услуг эффективно решать проблему утилизации вычислительных мощностей без ущерба конечным пользователям. В результате разница такой утилизации на уровне пользователя (около 30 %) и на уровне поставщика (до 85 %) делает облачную модель услуг экономически привлекательной. Более того, этот факт позволяет лояльно относиться к потерям на уровне виртуализации. Ввиду того, что при научных расчётах утилизация зачастую достигает 100 %, приведённое выше предположение не работает, и потери на виртуализацию могут оказаться тем фактом, который перечеркнёт все преимущества облачного подхода.

Коноплев Вениамин Викторович — главный специалист, кандидат технических наук, v.konoplev@cosmos.ru

Боярский Михаил Николаевич — главный специалист, mb@cosmos.ru

Назиров Равиль Равильевич — доктор технических наук, профессор, r.nazirov@cosmos.ru

Другое обстоятельство, которое следует учитывать при развёртывании вычислительных задач в облачной среде, это возможность интенсивной работы с общими наборами данных. Большинство облачных платформ ориентированы на изолированные разделы данных, которые оформлены как блочные устройства и привязаны к определённым виртуальным машинам. Такой подход неизбежно приводит к значительным потерям производительности при попытке доступа к локальному разделу из соседних вычислительных узлов.

В 2013–2014 гг. в ИКИ РАН был разработан и внедрён программно-аппаратный комплекс SCARP, который представляет собой отказоустойчивую и масштабируемую платформу виртуализации приложений с высокопроизводительным однородным доступом к общим массивам данных. Данный комплекс представляет собой технологическую связку вычислительного кластера на базе лёгкой, контейнерной виртуализации OpenVZ (OpenVZ Virtuozzo containers Wiki, URL: https://openvz.org/Main_Page, дата обращения: 23.11.2015), с высокопроизводительным отказоустойчивым масштабируемым хранилищем на базе параллельной файловой системы Lustre (<http://lustre.opensfs.org/>, дата обращения: 23.11.2015). В результате было создано экономически эффективное решение для виртуализации «тяжёлых» расчётных задач и предоставления облачных сервисов, в котором минимизированы потери на уровне виртуализации и решена проблема масштабируемости при разделяемом доступе к общим наборам данных.

АРХИТЕКТУРНЫЕ ОСОБЕННОСТИ КОМПЛЕКСА SCARP

Программно-аппаратный комплекс SCARP реализует симметричную архитектуру с аппаратной избыточностью, в которой отсутствуют единые точки отказа, а также узлы, агрегирующие потоки данных (рис. 1). Вычислительный кластер связан с параллельным хранилищем посредством высокоскоростной неблокируемой коммутационной сети с поддержкой технологии прямого удалённого доступа к памяти (RDMA) (Remote Direct Memory Access, URL: https://en.wikipedia.org/wiki/Remote_direct_memory_access, дата обращения: 23.11.2015). Параллельное хранилище представляет собой набор объектных серверов, которые обслуживают соответствующие дисковые пулы. Узлы вычислительного кластера работают непосредственно с объектными серверами, обслуживающими соответствующие блоки данных.

Параллельное файловое хранилище может обслуживать несколько параллельных файловых систем. При этом каждая файловая система распределена по всем объектным серверам пропорционально ёмкости их дисковых пулов. Таким образом, происходит условно-равномерное балансирование пользовательской нагрузки по всем дисковым пулам. Свободное пространство дисковых пулов может свободно перераспределяться между файловыми системами (в том числе в сторону уменьшения). На дисковых пулах предусмотрена возможность кэширования часто используемых данных на твердотельных дисках.

Логической единицей предоставления услуг в комплексе SCARP является *проект*. Он представляет собой набор виртуальных машин с согласованной политикой доступа, которые привязаны к выделенной параллельной файловой системе. Образы виртуальных машин также размещены в параллельном хранилище. Виртуальные машины могут свободно мигрировать между физическими серверами. Виртуальные машины реализуются как контейнеры OpenVZ. Контейнерная виртуализация позволяет виртуальной машине использовать прямой канал RDMA для доступа к данным параллельного хранилища без потерь на программном гипервизоре.

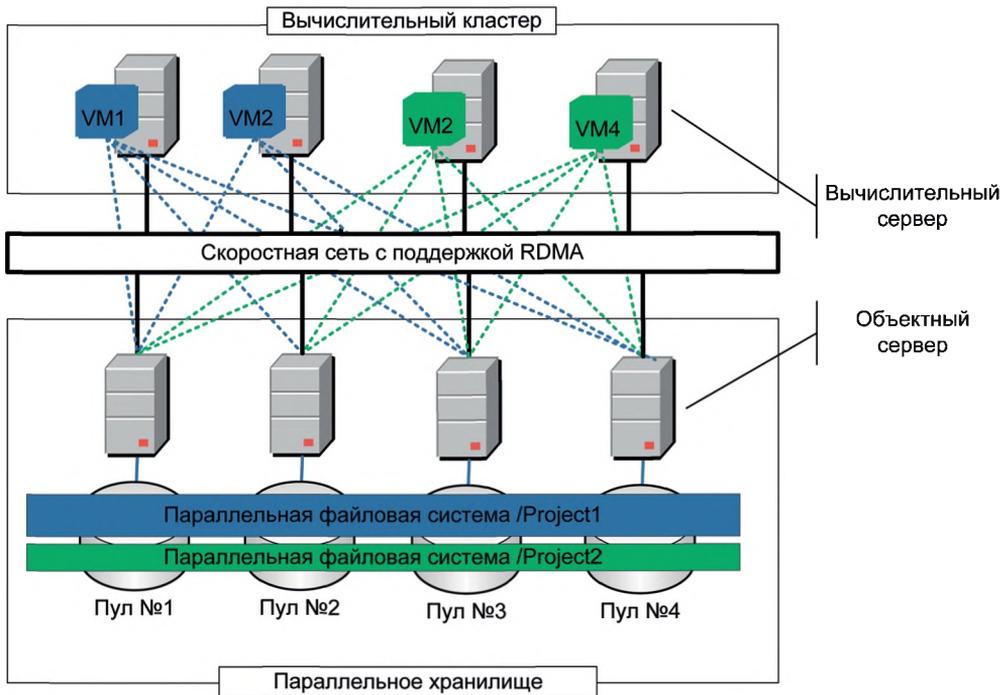


Рис. 1. Архитектурные особенности SCARP

В целом архитектура комплекса SCARP позволяет получить систему со следующими потребительскими характеристиками:

- масштабируемость до нескольких петабайт;
- производительность порядка нескольких гигабайт в секунду;
- единое пространство имён;
- отказоустойчивость с полным аппаратным дублированием;
- повышенная надёжность хранения данных;
- однородный высокопроизводительный доступ к данным со всех вычислительных узлов;
- работа в широком профиле нагрузок и сценариев, включая как последовательный, так и произвольный доступ к данным.

РЕАЛИЗАЦИЯ КОМПЛЕКСА SCARP

Выбор параллельной файловой системы

При подборе подходящей параллельной файловой системы для реализации параллельного хранилища рассматривались три наиболее популярных свободно распространяемых открытых программных решения:

- распределённая файловая система GlusterFS (Storage for your cloud, <http://www.gluster.org/>, дата обращения: 23.11.2015);
- параллельное многофункциональное хранилище CEPH (URL: <http://ceph.com/>, дата обращения: 23.11.2015);
- упомянутая выше параллельная файловая система Lustre;

Не углубляясь в подробный сравнительный анализ данных программных продуктов, можно привести причины, по которым выбор был сделан в пользу файловой системы Lustre:

- на момент проектирования только файловая система Lustre официально поддерживала сети с механизмом RDMA;
- на момент проектирования комплекса SCARP файловый режим работы в CEPH не был готов для использования в штатном (production) режиме;
- на данный момент файловая система Lustre является единственным из трех вариантов, который оптимизирован под работу с окончательным хранилищем на базе ZFS (<http://ru.wikipedia.org/wiki/ZFS>, дата обращения: 23.11.2015);
- в силу архитектурных особенностей файловая система Lustre имеет наилучшие показатели в скорости работы с метаданными;
- файловая система Lustre имеет наилучшую репутацию, как продукт, активно используемый в суперкомпьютерах из списка TOP500.

Общая функциональная схема комплекса SCARP

Ниже на рис. 2 показана логическая схема организации программно-аппаратного комплекса SCARP, реализованная в ЦОД (Центре обработки данных) ИКИ РАН. Она включает модуль данных, модуль метаданных, вычислительный кластер, контроллер домена, шлюз для внешних взаимодействий и высокоскоростную неблокируемую внутреннюю сеть передачи на базе технологии Infiniband QDR с пропускной способностью 40 Гбит/с. Архитектура предусматривает полное аппаратное дублирование компонент комплекса, включая сетевые соединения.

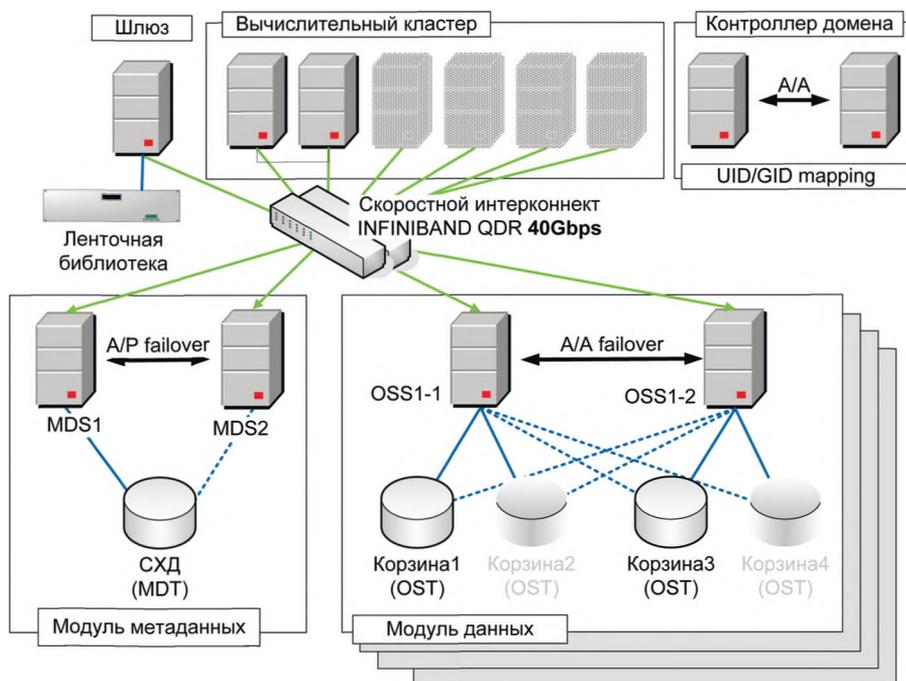


Рис. 2. Логическая организация программно-аппаратного комплекса

Вычислительный кластер состоит из четырёхпроцессорных серверов. Это позволяет уменьшить количество узлов сети и, как следствие, увеличить предел масштабируемости, при котором весь комплекс может обслуживаться одним центральным коммутатором, что гарантирует неблокируемость коммутационной матрицы.

Модуль метаданных, обслуживающий карты размещения данных в параллельных файловых системах, состоит из двух серверов (MDS), работающих по схеме активный-пассивный с разделяемым хранилищем метаданных.

Базовая единица масштабируемости хранилища — модуль данных. Он состоит из двух объектных серверов (OSS), которые работают в связке активный-активный и могут обслуживать до четырёх внешних корзин ёмкостью по 60 дисков каждая.

Контроллер домена реализует контроль доступа к сервисам комплекса SCARP, отображение между символьными именами и числовыми идентификаторами пользователей и групп, а также базовые сетевые сервисы, включая DNS и DHCP. Он организован на базе RedHat Identity Management и работает на двух серверах по схеме активный-активный.

Особенности организации файловых систем в SCARP

В текущей аппаратной реализации модуль данных может обслуживать до четырёх 60-дисковых корзин. Каждая дисковая корзина соответствует отдельному пулу ZFS (Zettabyte File System) (ZPOOL). Использование файловой системы ZFS в качестве окончательного хранилища дает возможность легко создавать большое количество объектных томов на базе дискового пула и гибко перераспределять дисковое пространство между ними. В частности это позволяет реализовать заявленную схему равномерного распределения параллельной файловой системы по общему дисковому пространству параллельного хранилища и выполнить балансирование нагрузки. При организации параллельной файловой системы на всех пулах ZFS создаются объектные тома (OST), которые потом комбинируются в распределённую файловую систему Lustre, как показано на рис. 3.

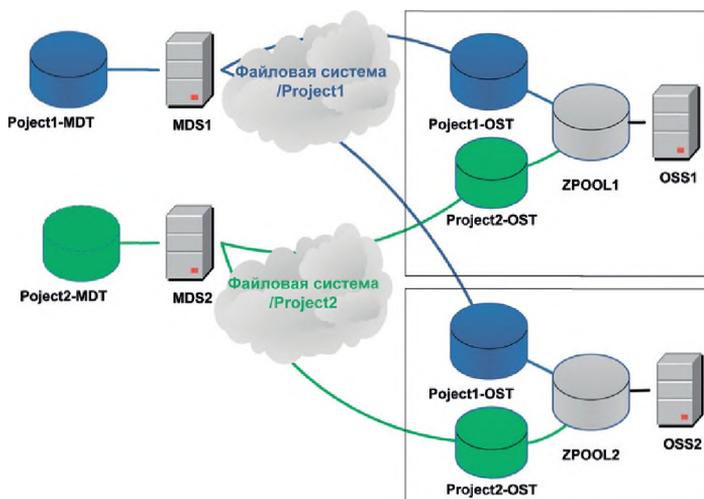


Рис. 3. Особенности организации параллельной ФС в комплексе SCARP

Защита данных

Рассматриваемый программно-аппаратный комплекс предусматривает защиту данных при двух видах негативного воздействия:

- выход из строя жёстких дисков в дисковых пулах;
- потеря (удаление или искажение) данных в файловой системе Lustre.

Для защиты от выхода из строя жестких дисков используются группы с тройной чётностью RAID-Z3, состоящие из 19 дисков объёмом по 4 ТБ каждый. Оценочные расчёты показывают, что при такой организации хранения вероятность потери данных в хранилище объёмом 1 ПБ в течение 10 лет находится в пределах одного процента.

Потеря данных может быть следствием ошибок в коде или конфигурации файловой системы Lustre, а также ошибок пользователя либо пользовательского программного обеспечения. Для защиты от потерь данного типа используется механизм «снимков» в оконечных объектных хранилищах на базе ZFS. Со всех объектных томов в параллельном хранилище делаются ежедневные снимки. Политика хранения снимков может настраиваться индивидуально для каждой параллельной файловой системы. При необходимости существует возможность сделать откат к одной из предыдущих версий файловой системы, либо собрать «исторический клон» файловой системы из снимков, сделанных на определённую дату, и восстановить из него потерянные данные.

Схема доступа и производительность

Условная схема подключения хранилища для большого проекта, работающего под управлением OS (операционная система) Linux, показана на рис. 4. В процессе работы комплекса все параллельные файловые системы хранилища смонтированы на всех физических вычислительных узлах. При запуске виртуальной машины OpenVZ в неё экспортируется ветка соответствующей параллельной файловой системы через механизм связанного монтирования (bind mount). Поскольку контейнер виртуальной машины использует ядро физического вычислительного узла, скорость ввода-вывода для неё практически соответствует скорости ввода-вывода для физического узла.

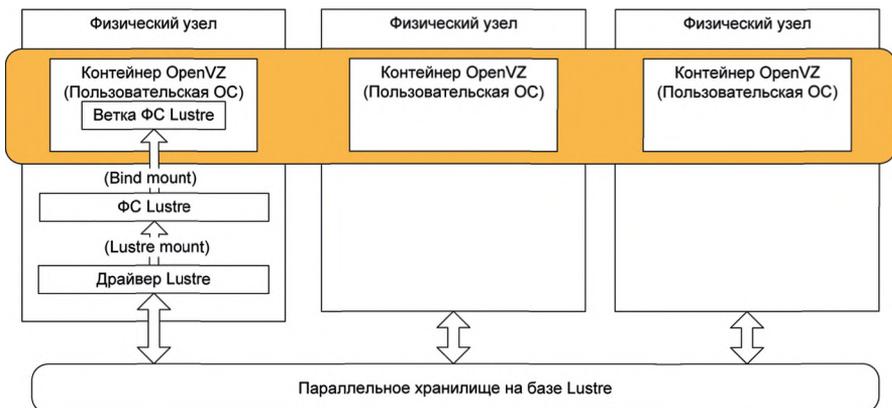


Рис. 4. Схема доступа для контейнеров OpenVZ к параллельному хранилищу

Ниже в таблице показаны результаты тестов производительности для системы из двух вычислительных узлов, двух объектных серверов и двух дисковых корзин. Как видно, данные, полученные для одного потока, соизмеримы со скоростью работы локальных дисковых хранилищ. В тоже время интегральная скорость последовательного ввода-вывода в комплексе может достигать нескольких гигабайт в секунду.

Производительность доступа к хранилищу из контейнера

	Для одного потока	Интегрально
Запись [МБ/с]	234	~1500
Перезапись [МБ/с]	166	н/д
Чтение [МБ/с]	413	~2000
Создание [1/с]	318	2079
Удаление [1/с]	2197	3387

Текущие недостатки и ограничения

При создании и эксплуатации комплекса SCARP были выявлены следующие ограничения:

- текущая реализация ФС Lustre не поддерживает размещение синхронного журнала записи (SLOG) на выделенных устройствах в оконечном хранилище ZFS. Таким образом, функция кэширования записи на дисках SSD пока недоступна;
- текущая реализация ФС Lustre пока не поддерживает монтирование отдельных веток файловой системы. На физическом узле можно монтировать файловые системы только от корня. Этот недостаток компенсируется на стадии реэкспорта отдельной ветки файловой системы внутрь контейнера;
- текущая реализация ФС Lustre имеет ограничения по обнаружению ошибок целостности при использовании оконечного хранилища на базе ZFS. Недостающая функциональность должна быть добавлена в следующей стабильной версии 2.5.4;
- ввиду ограничений системы виртуализации OpenVZ, приложениям внутри контейнера недоступен механизм RDMA. Это создаёт ограничения при использовании механизма MPI. Приложения, активно использующие MPI, необходимо запускать непосредственно на физических узлах. Для этого можно использовать отдельный вычислительный кластер, подключённый к параллельному хранилищу;
- текущая версия OpenVZ поддерживает размещение файлов-образов виртуальных машин только на файловых системах EXT3/4 и NFS. Для размещения файлов-образов в параллельном хранилище в настоящее время применяется реэкспорт через сервер NFS локального физического узла.

ТЕКУЩИЕ ПРИЛОЖЕНИЯ КОМПЛЕКСА SCARP

В настоящее время комплекс SCARP обслуживает три научных проекта и одно корпоративное облачное приложение, а именно:

- архив данных дистанционного зондирования Земли из космоса и набор картографических приложений для работы с ними;
- архив и приложения для работы с данными астрофизических рентгеновских телескопов;
- архив проекта ВЕНЕРА-ЭКСПРЕСС и научные инструменты для планетарных исследований;
- корпоративное облачное хранилище SeaFile (URL: <http://ru.wikipedia.org/wiki/ZFS>, дата обращения: 23.11.2015), интегрированное с корпоративным сервером совместной работы Kolab (URL: <https://www.seafile.com/>, дата обращения 23.11.2015).

В целом, несмотря на некоторые недостатки и ограничения, изложенные в предыдущей главе, выбранная архитектура комплекса SCARP показала вполне обнадеживающие результаты и это направление, несомненно, будет развиваться в дальнейшем в центре обработки данных ИКИ РАН.

SCARP: SCALABLE AND FAULT TOLERANT LIGHT-WEIGHT VIRTUALIZATION PLATFORM FOR DATA INTENSIVE AND CLOUD APPLICATIONS

V. V. Konoplev, M. N. Boyarskiy, R. R. Nazirov

Space Research Institute of Russian Academy of Sciences (IKI RAN)

This paper represents the architecture and the deployment of the scalable and fault tolerant computing and storage system SCARP. It was created in the Space Research Institute of RAS for virtualization of data intensive scientific applications which require shared collaboration data access. The SCARP system uses open software and commodity hardware components allowing the project to keep cost effectiveness. Lightweight virtualization technologies made possible a highly customizable operating system environment to be combined with a near-bare-metal performance. Particularly the testing results revealed I/O performance between the virtual machine and the parallel shared storage turned out to be comparable to internal I/O performance of a dedicated physical server. The represented system can be leveraged as a technological platform for cloud services providing both scientific and corporate applications with shared and intensive data access needs.

Keywords: virtualization, cloud computing, OpenVZ, Lustre, ZFS, parallel file systems.

Konoplev Veniamin Vitorovich — main specialist, PhD, v.konoplev@cosmos.ru

Boyarskiy Mikhail Nikolaevich — main specialist, mb@cosmos.ru

Nazirov Ravil Ravilievich — doctor of technical sciences, professor, mazirov@cosmos.ru

ОПТИМИЗАЦИЯ АЛГОРИТМА BACKFILL И СИСТЕМЫ ПЛАНИРОВАНИЯ ЗАДАНИЙ ДЛЯ ИСПОЛЬЗОВАНИЯ НА СУПЕРКОМПЬЮТЕРЕ «ЛОМОНОСОВ»

С. Н. Леоненков^{1,2}, С. А. Жуматий²

¹ Московский государственный университет имени М. В. Ломоносова (МГУ)

² Научно-исследовательский вычислительный центр МГУ

Менеджер ресурсов SLURM, используемый на многих ведущих суперкомпьютерах (СК), в том числе на СК «Ломоносов», имеет встроенный планировщик и систему учёта и лимитов. Однако реальный опыт применения SLURM на суперкомпьютере в режиме коллективного пользования показал, что встроенных возможностей не хватает для оптимальной загрузки. Авторами был разработан внешний планировщик, подключаемый к SLURM через интерфейс `wiki2`, реализующий оптимизированный алгоритм Backfill и собственный набор лимитов. В работе представляется развитие алгоритмов планирования, реализуемых в планировщике, в частности алгоритм «скользящих» приоритетов, позволяющих менять скорость продвижения задач в очереди.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 13-07-12206).

Ключевые слова: SLURM; алгоритмы планирования; Backfill; система приоритетов.

ВВЕДЕНИЕ

Каждый год вместе с производительностью суперкомпьютерных комплексов по всему миру не менее быстро растёт и база пользователей подобных систем. Ресурсы этих систем становятся все более востребованными для решения научных, промышленных, финансовых и прочих задач разной сложности. Пользователи высокопроизводительных кластеров имеют различные права доступа, запрашивают от одного до многих тысяч процессоров для запуска своих задач, время выполнения которых колеблется от нескольких секунд до многих часов или даже дней. Удовлетворение всех запросов пользователей в режиме реального времени в рамках ограниченных мощностей становится краеугольным камнем для системных администраторов загруженных вычислительных систем.

Суперкомпьютер «Ломоносов», используемый для исследований, которые проводят учёные и студенты МГУ, занимает второе место в рейтинге Top50 суперкомпьютеров, установленных на территории СНГ и показавших наибольшую производительность на тесте Linpack (по состоянию на 31 марта 2015 г.) (<http://top50.supercomputers.ru/?page=rating>), и удерживает 94-е место в авторитетном международном рейтинге самых высокопроизводительных вычислительных систем Top500 (по состоянию на ноябрь 2015 г.). (<http://www.top500.org/lists/2014/11/>).

Система состоит из 12 422 процессоров [Sadovnichy et al., 2013], но даже такого количества недостаточно, чтобы обеспечить все группы исследователей, работающих на суперкомпьютере «Ломоносов». Все процессоры физически раз-

Леоненков Сергей Николаевич — аспирант, leonenkovs@parallel.ru

Жуматий Сергей Анатольевич — ведущий научный сотрудник, кандидат физико-математических наук, serg@parallel.ru

делены на группы по 4 или 6 штук, называемые узлами, которые логически объединены в разделы. Каждый раздел ориентирован на определённые потребности пользователей (<http://www.top500.org/system/177421>).

В текущей конфигурации «Ломоносова» доступно 8 разделов, количество узлов в каждом колеблется от 1 до 4096. В среднем в день выполняется от 200 до 400 задач, которые ставят на запуск 30–50 разных пользователей, причём каждый год активно используется около 400 аккаунтов, и эти числа не являются пределом.

Чтобы справляться с ежедневно растущими нагрузками на суперкомпьютер «Ломоносов», на нем используется система SLURM (Simple Linux Utility for Resource Management).

1. НАУЧНАЯ ОБЛАСТЬ И СОПУТСТВУЮЩИЕ ПОНЯТИЯ

SLURM — это высокомасштабируемый, отказоустойчивый менеджер кластеров и планировщик заданий для больших систем вычислительных узлов. Lawrence Livermore National Laboratory начала разработку SLURM в 2001 году для управления ресурсами собственных суперкомпьютеров (<http://slurm.schedmd.com/slurm.html>) [Jette, Grondona, 2003]. Менеджер кластеров SLURM используется на многих суперкомпьютерах из рейтинга Top500, включая половину суперкомпьютеров из первой десятки рейтинга (по состоянию на ноябрь 2013 г.). Исходный код написан на языке программирования C. Система доступна по лицензии GNU GPL V2 и хорошо задокументирована. Две главные цели, которые ставили перед собой разработчики, — высокая масштабируемость и портируемость SLURM.

В основе SLURM реализована иерархическая архитектура управления суперкомпьютером. Главный агент менеджера — основной контроллер, который содержит управляющий агент `slurmctld`. В некоторых архитектурах для повышения отказоустойчивости менеджера основному контроллеру может соответствовать резервный. Основная задача `slurmctld` — выделение ресурсов под поставленные пользователями задачи.

Для управления каждым отдельным узлом на них развернуты демоны `slurmd`. Они служат для запуска и управления заданием на самом узле (получение задания от основного контроллера, запуск уже непосредственно на ядрах) и мониторинга его состояния.

SLURM спроектирован для гетерогенных кластеров с допустимым количеством процессоров до 10 миллионов штук. SLURM успешно апробирован и используется на суперкомпьютере с более чем 98 000 вычислительными узлами. Пользователи суперкомпьютера, управляемого с помощью SLURM, могут ставить на исполнение 1000 задач в секунду. Менеджер может исполнять до 500 задач в секунду (в зависимости от конфигурации и оборудования системы).

Пользователи SLURM имеют большую свободу в установке настроек для каждой отдельной задачи. Можно изменять в том числе её размер и лимит по времени выполнения.

Одно из основных преимуществ SLURM — модульность: ему доступны десятки дополнительных плагинов. Написаны стандартные планировщики, один из которых (`sched/backfill`) используется для оптимизации работы суперкомпьютера «Ломоносова». При возникновении ситуации, когда стандартные алгоритмы планирования полностью не устраивают системного администратора, можно легко ввести в эксплуатацию собственный плагин, используя стандартные интерфейсы (`wiki` и `wiki2`), которые предоставляет SLURM.

Системный администратор может достаточно гибко настраивать логическую конфигурацию вычислительной системы, которую будет поддерживать SLURM, и легко варьировать множество параметров кластера путём изменения конфигурационного файла или использования соответствующих команд.

Специфика использования менеджера на суперкомпьютере «Ломоносов» предполагает расширение функциональности программного комплекса SLURM для качественного администрирования и повышения эффективности системы.

1.1. Алгоритм Backfill

На суперкомпьютере «Ломоносов» используется стандартный планировщик и алгоритм Backfill. Этот алгоритм планирования используется на большом количестве кластеров и доказал свою высокую эффективность. Исследования показали, что он позволяет повысить плотность использования ресурсов суперкомпьютеров примерно на 20 % и уменьшить среднее время ожидания постановки задач на исполнение [Jette, Grondona, 2003].

В основе алгоритма Backfill лежит принцип First Come First Served. Это означает, что задачи запускаются в том же порядке, в котором они были поставлены в очередь (при доступности ресурсов для исполнения каждой конкретной задачи). В процессе работы алгоритма Backfill обрабатывается очередь задач, ожидающих запуска на выполнение. Задачи в очереди упорядочены по их приоритету и времени постановки в очередь. Как только система SLURM сигнализирует, что появилась новая задача или одна из задач закончила выполнение и освободила ресурсы, то алгоритм (при наличии необходимого количества свободных узлов для запуска первой в очереди задачи) выделяет эту задачу в очереди и запускает её на исполнение. В противном случае, если для запуска первой в очереди задачи свободных узлов не хватает, то планировщик пытается совершить «уплотнение» очереди: взять следующую в очереди задачу и поставить её на выполнение при условии, что для запуска имеется необходимое количество свободных узлов и не нарушатся права остальных задач в очереди на запуск (т. е. их запуск не будет задержан). Примеры работы алгоритмов FCFS и FCFS + Backfill представлены на рис. 1 [Baraglia et al., 2008].

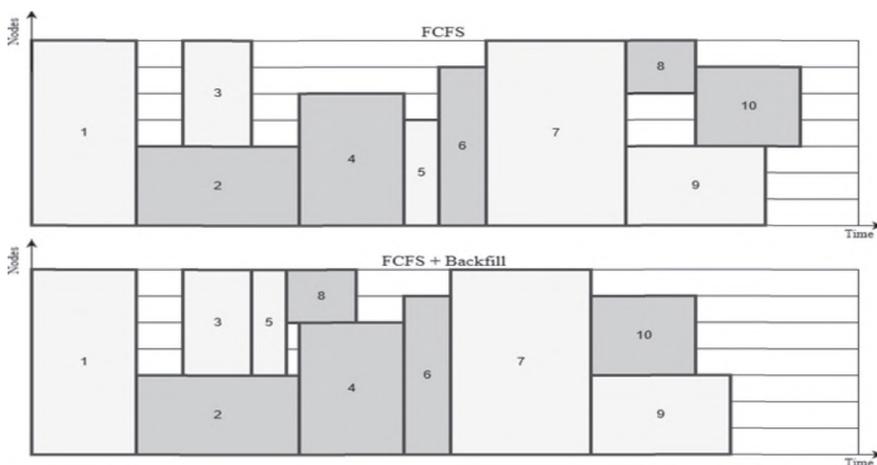


Рис. 1. Примеры FCFS и FCFS + Backfill

1.2. Система приоритетов

Вычисление приоритетов в архитектуре системы SLURM на суперкомпьютере «Ломоносов» отдано специальному плагину, который идёт в стандартной комплектации менеджера ресурсов, Multifactor Priority Plugin. Плагин реализует систему динамических приоритетов. Приоритет задания зависит от пяти факторов (в том числе: время ожидания задачи, размер, партиция), у каждого из которых есть вес, заданный системным администратором при настройке системы. Далее вычисляется взвешенная сумма факторов, которая и назначается каждой задаче, как её приоритет. Формула приоритета каждой конкретной задачи представляет собой взвешенную сумму факторов, влияющих на приоритет (см. ниже).

Формула для вычисления значения приоритета

$$\begin{aligned} \text{Job_priority} = & (\text{PriorityWeightAge}) * (\text{age_factor}) + \\ & (\text{PriorityWeightFairshare}) * (\text{fairshare_factor}) + \\ & (\text{PriorityWeightJobSize}) * (\text{job_size_factor}) + \\ & (\text{PriorityWeightPartition}) * (\text{partition_factor}) + \\ & (\text{PriorityWeightQOS}) * (\text{QOS_factor}); \end{aligned}$$

Далее приведены значения основных констант, которые используются в текущей конфигурации суперкомпьютера «Ломоносов».

Настройки Multifactor Priority Plugin

```
PriorityType=priority/multifactor
PriorityDecayHalfLife=14-0
PriorityWeightFairshare=0
PriorityWeightAge=10000 # default 1000
PriorityWeightPartition=10000000 # default 10000
PriorityWeightJobSize=0 # default 1000
PriorityMaxAge=3-0
```

Стандартный цикл работы планировщика на примере одной партиции выглядит, как представлено ниже (http://slurm.schedmd.com/slurm_ug_2012/SUG-2012-Scheduling.pdf).

Псевдокод стандартного цикла планировщика

- 1) создание очереди заданий для партиции;
- 2) сортировка заданий по приоритету;
- 3) удаление из очереди задачи, которые не представляется возможным запустить;
- 4) нахождение узлов для запуска заданий;
- 5) «уплотнение», если требуется;
- 6) запуск заданий;
- 7) обновление очереди и статусов задач и узлов;
- 8) возврат к п. 2.

1.3. Опыт использования SLURM на суперкомпьютере «Ломоносов»

Несмотря на все плюсы SLURM системные администраторы НИВЦ МГУ столкнулись с некоторыми проблемами при использовании системы на суперкомпьютере «Ломоносов», который обладает базой из примерно 1000 различных

активных пользователей (за годовой период). В каждый момент времени на суперкомпьютере «Ломоносов» работают более 20 уникальных пользователей. Типичная ситуация на суперкомпьютере «Ломоносов» — примерно 100 исполняемых задач и 200 задач в очереди в каждый момент времени. Из-за такой приличной нагрузки НИВЦ МГУ столкнулся с проблемой ограниченности набора настроек SLURM.

Одной из серьёзных проблем планирования процессорного времени является тот факт, что каждый пользователь ждёт слишком долго, чтобы начать свою работу даже при использовании планировщика, который основан на алгоритме Backfill. Это вызывает очень распространённую ситуацию: каждый пользователь ставит в очередь несколько заданий (в осязаемом количестве случаев с целью отладки работы своей программы) и ждёт. Когда первая задача пользователя запускается, верхняя часть очереди заполнена другими его задачами, так что большинство пользователей ожидают одного. Эта проблема не может быть решена путём установки ограничения на максимальное количество выполняемых задач пользователя — эффективность работы значительно снижается, когда очередь содержит много коротких и/или небольших задач.

Мы предлагаем новый тип лимитов - лимит процессоро-часов для каждого пользователя. (Процессоро-часы рассчитываются как время, на которое процессоры было выделены на конкретные задачи одного пользователя, умноженное на их количество.) Это позволяет управлять количеством процессоро-часов (или ядро-часов, или узло-часов) всех исполняемых задач и задач для каждого пользователя, которые уже стоят в очереди. Таким образом, каждый пользователь может запускать, например, одну задачу для большого количества узлов на короткое время, или больше задач на меньшее количество процессоров.

Другой недостаток SLURM — неспособность установить большинство ограничений для разделов по умолчанию. Например, максимальное количество запущенных заданий может быть установлено только для всего кластера на одного пользователя. Если у нас есть около 1000 пользователей, значение лимитов (так называемый Account в SLURM) для каждого пользователя добавляется в течение 1 минуты. Во время этой операции клиентские команды (sbatch, sinfo и т. д.) не могут быть выполнены. Это значит, что встроенная система лимитов SLURM полностью непригодна для наших нужд, и мы должны реализовать её внешними плагинами.

Для повышения эффективности использования суперкомпьютера «Ломоносов» мы должны добавить в планировщик некоторые другие функции. В целях повышения прозрачности системы приоритетов было решено отказаться от плагина Multifactor и добавить новую систему приоритетов во внешний планировщик. Новая система основана на расслоении пользователей — администратор организует систему приоритетов задач разделением на уровни (от 0 до 65536 в нашем случае), формируя группы из высокого и низкого приоритета пользователей. Ежегодный опрос пользователей показал себя очень эффективным инструментом, используя который системные администраторы могут, например, поставить недобросовестным пользователям уменьшенный приоритет.

Для оптимизации работы планировщика SLURM было решено внести следующие дополнения:

- учёт и контроль процессоро-часов, запрошенных каждым пользователем;
- создание прозрачной системы приоритетов с возможностью настройки в режиме реального времени;
- реализация возможности использования узлов из разных разделов для пользователей с определённым приоритетом;

- добавление квот по времени: определённое число процессоро-часов в неделю/месяц/год;
- реализация возможности использования разных алгоритмов планирования в разных очередях.

Все эти дополнения не могут быть реализованы без расширения функциональности SLURM, SLURM поддерживает интерфейсы для внешних планировщиков. Есть несколько коммерческих планировщиков (например, MOAB, URL: <http://www.adaptivecomputing.com/products/hpc-products/moab-hpc-suite-enterprise-edition/>), которые могут быть использованы совместно со SLURM. MOAB имеет очень широкие возможности, но является коммерческим приложением. Планировщик MAUI (<http://www.adaptivecomputing.com/support/download-center/maui-cluster-scheduler/>) так же поддерживается SLURM, но он больше не разрабатывается и через некоторое время будет закрыт компанией-владельцем, которая работает сейчас над MOAB). По этой причине использование MAUI не безопасно, и, кроме того, в нём отсутствуют некоторые необходимые функции.

Что касается opensource-планировщиков для SLURM, мы нашли только два проекта, которые пытаются реализовать плагины SLURM. Первый проект называется “slurm-spank-plugin” (<https://code.google.com/p/slurm-spank-plugins/w/list>), и он не развивается с 2011 г. Проект состоит из пяти различных плагинов для дополнения функций SLURM.

Второй проект называется IPSCHEM (<http://code.google.com/p/slurm-ip-sched/>), он направлен на поддержку новых функций планировщика SLURM (ориентирован на взаимодействие CPU и GPU), в нём также отсутствуют необходимые нам функции.

Больше мы не нашли каких-либо других упоминаний или публикаций о внешних планировщиках для SLURM. В итоге было решено реализовать свой собственный планировщик, чтобы наиболее эффективно использовать вычислительные возможности суперкомпьютера «Ломоносов» с его характерной высокой рабочей нагрузкой и постоянно растущей пользовательской базой.

2. ВНЕШНИЙ ПЛАНИРОВЩИК ДЛЯ SLURM

Согласно требованиям к программному обеспечению (ПО), предъявляемым в Суперкомпьютерном центре МГУ, на техническое решение были наложены некоторые ограничения. Периодически команда разработчиков SLURM выпускает новую версию менеджера, и поэтому наши нововведения должны быть переносимыми. Все новые функции, которые были внесены в программный код SLURM, должны быть за разумное время переносимы в новую версию менеджера. Это требование основное, из него вытекает множество побочных ограничений, с которыми пришлось столкнуться. Например, изменения не могут вноситься в ядро системы — фоновые программы `slurmd` и `slurmctld`, потому что это крайне затормозит процесс обновления. Намного практичнее изменять внешние модули системы. Направленность ядра SLURM на внешние плагины не позволяет создавать свои плагины, не ориентируясь на интерфейсы `wiki` и `wiki2` (<http://code.google.com/p/slurm-ipsched/>).

Другим затрудняющим разработку фактором стало отсутствие возможности добавления новых «состояний» и их описаний без затрагивания ядра SLURM. Основной цикл планировщика описан в файле `job_scheduler.c`, который является частью утилиты `slurmctld`. Если все же добавить новые «состояния» и их описания

во внешних файлах, то они будут неправильно восприняты или проигнорированы `slurmctld`. Последней серьёзной проблемой стало отсутствие возможности использования встроенных методов хранения настроек новых дополнений. Встаёт необходимость создания своего хранилища данных, применяющихся для корректной работы новых дополнений.

Таким образом, для реализации всех дополнений самым оптимальным решением представлялось создание внешнего планировщика, который будет использовать стандартный интерфейс `wiki2`.

Стандартные интерфейсы `wiki/wiki2` были изначально разработаны для поддержки внешних планировщиков: `Maui Cluster Scheduler™` и `Moab® HPC Suite`. Внешне они представляют с собой подобие плагинов самого штатного планировщика, и даже использует те же примитивы и функции, но отличие состоит в том, что вся логика планировщика полностью отсутствует и оставлена в компетенции внешнего модуля, который и отвечает за алгоритмы планирования [Jackson et al., 2007].

Алгоритм работы `wiki2` завязан на обмене сообщениями с внешним планировщиком. Между `wiki2` и внешним планировщиком налажены два канала связи: первый служит для передачи команд и получения ответов, второй — для информирования внешнего планировщика о событиях, произошедших в системе.

Базовый набор команд внешнего планировщика: `CANCELJOB`, `GETJOBS`, `JOBMODIFY`, `NOTIFYJOB`, `STARTJOB`, `INITIALIZE`, `REQUEUEJOB`, `SUSPENDJOB`, `RESUMEJOB`, `SIGNALJOB`. Каждая из них запрашивает какую-либо информацию от `wiki2` или передаёт указание на изменение состояния конкретной задачи. В канале «событий» внешний планировщик может получить лишь два типа сообщения: «1234» или «1235». Первое соответствует изменению параметра задания, второе — изменению в структуре/конфигурации `SLURM`.

Данная архитектура (рис. 2) позволяет в полной мере воссоздать функционал штатного планировщика и дополнить его новыми свойствами. Основная реализация дополнения (написание нового внешнего планировщика) представляет собой модуль, содержащий две части: коммуникационный менеджер и сам внешний планировщик, который будет нести в себе всю логику алгоритма планирования [Novotny, 2009].

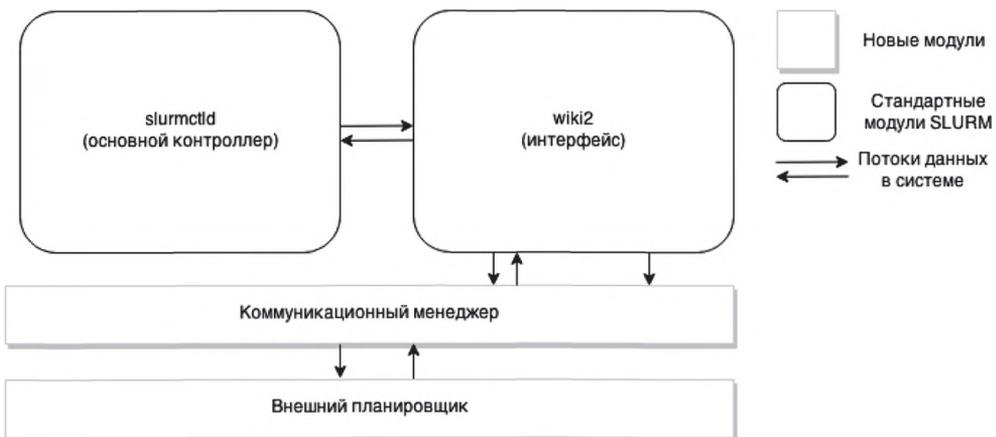


Рис. 2. Архитектура решения

3. ЭФФЕКТИВНОСТЬ РЕАЛИЗОВАННОГО РЕШЕНИЯ

Предложенное нами решение было протестировано на исторических данных очереди regular4 (самая большая очередь суперкомпьютера «Ломоносов»). Лимит процессоро-часов был выбран так, чтобы позволить пользователям комфортно использовать возможности системы, но в то же время ограничить возможность отдельному пользователю занять часть кластера почти полностью. Например, для очереди regular4 лимит процессоро-часов равен 81 920. Этот предел позволил перераспределить до 2,7 % от месячного процессорного времени на работу других пользователей, что увеличило количество ресурсообеспеченных пользователей.

Кроме того, внешний модуль с новым планировщиком был протестирован в сравнении со стандартным плагином-планировщиком SLURM (на исторических данных в очереди regular4 суперкомпьютера «Ломоносов»). Он показал хорошее ускорение для времени старта «первой» задачи каждого пользователя (запуск до ~9 % быстрее по сравнению со стандартным плагином), в то время как общая эффективность очереди почти не изменилась (только 0,2 % по сравнению со стандартным плагином). Сравнение работы SLURM со стандартным плагином Backfill и с нашим внешним планировщиком показано в таблице. Для тестирования мы использовали опцию SLURM, которая позволяет эмулировать большой кластер. Набор задач для тестов — реальный набор задач, которые были запущены в период с октября (2014) по декабрь (2014) в очереди regular4 суперкомпьютера «Ломоносов». Почти без падения общей производительности планирования мы добились ощутимых результатов в ускорении времени начала первых задач каждого пользователя.

Тесты на реальных данных

Эксперименты	Набор данных (ускорения)	
	Задержка старта (общая)	Задержка старта (первой задачи)
Backfill	0,998	1
Backfill + Лимит	1	0,907

4. НАПРАВЛЕНИЕ ДАЛЬНЕЙШИХ РАБОТ

Предлагаемое решение было разработано для проблем, с которыми мы столкнулись во время эксплуатации SLURM. Оно может эксплуатироваться любыми группами пользователей, использующих SLURM для управления суперкомпьютером. Решение имеет открытый исходный код и может быть быстро адаптировано для каких-либо других требований. Так же оно имеет хорошую портируемость благодаря использованию протокола wiki2. С помощью нашего планировщика эффективность суперкомпьютерного комплекса может быть повышена. Наши нововведения делают управление суперкомпьютером более гибким. Сейчас работа продолжается, и мы нацелены на расширение логики планирования в направлении «плавающих» приоритетов и локальных «ускорений» задач.

Механизм «ускорений» предполагает, что каждая задача будет обладать своим ускорением, счётчиком текущего ускорения и опцией “boost”. Ускорение будет накапливаться с каждым новым циклом планирования. Это позволит задачам продвигаться вверх в очереди с течением времени, даже если приоритет задач данного пользователя минимальный. Опция «boost» будет отвечать за стартовое количество позиций, на которое задача повысится в очереди, если её запуск будет

выполнен со специальным ключом. Данная опция будет лимитированной. Общий цикл планирования будет выглядеть следующим образом:

- 1) создание очереди заданий для партиции;
- 2) сортировка заданий по приоритету;
- 3) удаление задач, которые не представляется возможным запустить (CPU-limit);
- 4) загрузка данных из таблицы «ускорений»;
- 5) «ускорение» заданий из таблицы на указанное количество позиций;
- 6) нахождение узлов для запуска заданий;
- 7) «уплотнение», если требуется;
- 8) запуск заданий;
- 9) обновление очереди, статусов задач и узлов, таблицы «ускорений»;
- 10) возврат к п. 2.

Данные нововведения должны помочь избавиться от грубого расслоения пользователей на группы по значению приоритета.

ЛИТЕРАТУРА

- [Baraglia et al., 2008] *Baraglia R., Capannini G., Pasquali M., Puppini D., Ricci L., Techiouba A. D.* Backfilling Strategies for Scheduling Streams of Jobs on Computational Farms // Making Grids Work, II. 2008. P. 103–115. URL: http://link.springer.com/chapter/10.1007/978-0-387-78448-9_8.
- [Jackson et al., 2007] *Jackson D., Snell Q., Clement M.* Core Algorithms of the Maui Scheduler. Brigham Young University, Provo, Utah. Berlin Heidelberg, SSPP 2001, LNCS 2221. 2007. P. 87–102. URL: <http://www.eecs.harvard.edu/~chaki/bib/papers/jackson-01maui.pdf>.
- [Jette, Grondona, 2003] *Jette M., Grondona M.* SLURM: Simple Linux Utility for Resource Management // Proc. ClusterWorld Conference and Expo. San-Jose, California. 2003. 31 p.
- [Jones, 2012] *Jones M.* Optimization of resource management using supercomputers SLURM. 2012. URL: <http://www.ibm.com/developerworks/ru/library/l-slurm-utility/>.
- [Lipari, 2012] *Lipari D.* The SLURM Scheduler Design. URL: http://slurm.schedmd.com/slurm_ug_2012/SUG-2012-Scheduling.pdf.
- [Novotny, 2009] *Novotny M.* Job scheduling with the SLURM resource manager. Brno, 2009. 53 p. URL: https://is.muni.cz/th/173052/fi_b_b1/thesis.pdf.
- [Sadovnichy et al., 2013] *Sadovnichy V., Tikhonravov A., Voevodin V., Opanasenko V.* “Lomonosov”: Supercomputing at Moscow State University // Contemporary High Performance Computing: From Petascale toward Exascale. Chapman and Hall/CRC Computational Science. Boca Raton, United States: Boca Raton, United States, 2013. P. 283–307.
- [Snell et al., 2000] *Snell Q., Clement M. J., Jackson D. B., Gregory C.* The performance impact of advance reservation meta-scheduling // Job Scheduling Strategies for Parallel Proc., Lecture Notes in Computer Science. 2000. V. 1911. P. 137–153.

BACKFILL SCHEDULING SYSTEM OPTIMIZATION FOR USE ON LOMONOSOV SUPERCOMPUTER

S. N. Leonenkov^{1,2}, *S. A. Zhumatiy*²

¹ Research Computing Center, Lomonosov Moscow State University

² Lomonosov Moscow State University

The work proposes a design for a new external scheduler for SLURM (Simple Linux Utility for Resource Management). Schedulers, included in SLURM by default are good enough for many sites, but big supercomputers serving many users meet limitations of standard SLURM schedulers. In this work we discover methods to break these limitations via implementing new portable SLURM scheduler. We address the problem of maximizing the number of users, whose requests are processed in each given moment of time, and decrease start time of user's first task. Our approach is based on standard backfill algorithm and includes additional features, such as simplification of SLURM priority system, replacing slow SQL-based accounting checks by faster ACL checks and upgrading cluster administration convenience. In present we are developing backfill algorithm with "sliding" priority system, that allows system administrator to vary jobs "speed" in queue.

Keywords: SLURM; scheduling algorithms; Backfill.

Leonenkov Sergei Nikolaevich — postgraduate student, leonenkovs@parallel.ru

Zhumatiy Sergey Anatolievich — leading scientist, PhD, serg@parallel.ru

ЧИСЛЕННОЕ ОПИСАНИЕ НЕРАВНОВЕСНОГО КРИТИЧЕСКОГО ПОВЕДЕНИЯ ДВУМЕРНОЙ СТРУКТУРНО НЕУПОРЯДОЧЕННОЙ ХУ-МОДЕЛИ МЕТОДАМИ МОНТЕ-КАРЛО

И. С. Попов, П. В. Прудников, В. В. Прудников

Омский государственный университет им. Ф. М. Достоевского (ОГУ)

В работе даны результаты проведённого численного моделирования неравновесной критической релаксации структурно неупорядоченной двумерной ХУ-модели во всей низкотемпературной фазе Березинского в широком диапазоне спиновой концентрации. Моделирование выполнялось из низкотемпературного неравновесного состояния методами Монте-Карло. Исследованы эффекты старения и нарушения флуктуационно-диссипативной теоремы.

Работа выполнена при поддержке Российского научного фонда (проект № 14-12-00562). Для проведения расчётов были использованы ресурсы суперкомпьютерного комплекса Московского государственного университета имени М. В. Ломоносова и Межведомственного суперкомпьютерного центра Российской академии наук.

Ключевые слова: двумерная ХУ-модель, неравновесное критическое поведение, фазовый переход Березинского-Костерлица-Таулесса, структурный беспорядок, эффекты старения, флуктуационно-диссипативное отношение, методы Монте-Карло.

Исследование неравновесной критической динамики — актуальная научная задача [Tauber, 2014]. Вблизи критической точки возникают аномально большие, долгоживущие и сильно взаимодействующие флуктуации основных термодинамических величин. При этом релаксационные процессы существенно замедляются, что проявляется в эффектах критического замедления. Системы в критической точке оказываются примерами систем с медленной динамикой [Berthier, Kurchan, 2013], изучение которых вызывает в последние годы значительный научный интерес. Связано это с наблюдаемыми в них эффектами старения и нарушения флуктуационно-диссипативной теоремы.

Старение — это эффект замедления релаксационных процессов при увеличении времени, прошедшего с момента приготовления образца, — его «возраста», или времени ожидания t_w [Berthier et al., 2001]. Проявляется старение через динамические зависимости двухвременных характеристик систем, таких как автокорреляционные функции $C(t, t_w)$ и функции отклика $R(t, t_w)$. В процессе старения данные функции зависят не от разности времён $t - t_w$, а от t и t_w в отдельности, т. е. происходит нарушение однородности процесса релаксации во времени. Старение сопровождается нарушением флуктуационно-диссипативной теоремы [Berthier et al., 2001], связывающей динамические зависимости автокорреляционной функции и функции отклика системы.

Неравновесные релаксационные процессы вызывают особый интерес в критическом поведении двумерной ХУ-модели, в которой осуществляется топологический фазовый переход Березинского-Костерлица-Таулесса [Березинский, 1970, 2007; Kosterlitz, Thouless, 1973] при температуре T_{BKT} . Фазовый переход связан

Попов Иван Сергеевич — аспирант, diphosgen@mail.ru

Прудников Павел Владимирович — доктор физико-математических наук, профессор, prudnikp@univer.omsk.su

Прудников Владимир Васильевич — заведующий кафедрой, доктор физико-математических наук, профессор, prudnikov@univer.omsk.su

с диссоциацией связанных пар вихрь-антивихрь в точке перехода. Особенностью системы заключается в её аномально сильной пространственной и временной корреляции состояний системы во всей низкотемпературной фазе $T < T_{BKT}$, характеризующейся степенным законом спада [Kosterlitz, 1974], в то время как для термодинамических фазовых переходов второго рода эффекты сильной корреляции осуществляются лишь вблизи критической точки. Это позволяет наблюдать медленную динамику двумерной XY-модели не только вблизи критической точки, а во всем низкотемпературном диапазоне $T \leq T_{BKT}$. Двумерная XY-модель используется для описания поведения и свойств целого ряда физических систем (см. обзор [Коршунов, 2006]).

Дальний порядок в двумерной XY-модели разрушается при всех конечных температурах поперечными флуктуациями спиновой плотности. Низкотемпературные свойства системы связаны с образованием в низкотемпературной фазе вихревых квазимолекул [Березинский, 2007] и существованием как вихревой подсистемы [Березинский, 2007] — вихрей и антивихрей, так и подсистемы областей квазидальнего порядка [Berche et al., 2003], в которых спины локально упорядочены, а возбуждения имеют характер спиновых волн.

Наличие дефектов структуры приводит к возникновению эффекта пиннинга вихревых возбуждений на дефектах [Peçeira et al., 2003], что существенно замедляет динамические процессы в системе. Квазидальний порядок сохраняется с введением структурного беспорядка [Berche et al., 2003].

Неравновесное критическое поведение в структурно неупорядоченной двумерной XY-модели обусловлено неравновесными процессами в вихревой подсистеме, спин-волновыми процессами и взаимодействием этих подсистем с дефектами структуры. В неравновесном состоянии системы происходят процессы «рождения», «уничтожения» и «аннигиляции» вихревых элементарных возбуждений. Эффекты пиннинга играют значительную роль в процессе неравновесной релаксации системы из высокотемпературного начального состояния с замораживанием в низкотемпературной фазе [Прудников, Прудников и др., 2015б]. Было показано, что неравновесное критическое огрубление областей квазидальнего порядка существенно замедляется с введением в систему дефектов структуры [Prudnikov, Popov, 2015].

Известно [Berthier et al., 2001], что особенности неравновесного поведения систем с медленной динамикой, в частности эффекты старения, связаны с двухвременными зависимостями автокорреляционной функции $C(t, t_w)$ и функции отклика системы $R(t, t_w)$, определяемых следующим образом:

$$\left. \begin{aligned} C(t, t_w) &= \frac{1}{V} \int d^d x \left[\langle \mathbf{S}(x, t) \mathbf{S}(x, t_w) \rangle - \langle \mathbf{S}(x, t) \rangle \langle \mathbf{S}(x, t_w) \rangle \right], \\ R(t, t_w) &= \frac{1}{V} \int d^d x \left. \frac{\delta \langle \mathbf{S}(x, t) \rangle}{\delta h(x, t_w)} \right|_{h=0}, \end{aligned} \right\} \quad (1)$$

где $\mathbf{S}(x, t)$ — локальное спиновое поле, а угловые и квадратные скобки обозначают процедуры статистического усреднения по начальным состояниям и по различным реализациям поля структурного беспорядка соответственно. Временная переменная t_w характеризует время начала измерения (возраст образца) и называется временем ожидания. В неравновесных процессах с эффектами старения данные динамические функции зависят не только от времени наблюдения $t - t_w$, а от $t - t_w$ и t_w по отдельности, что имеет место при условии $t_w < t \ll t_{rel}$, где t_{rel} — время релаксации системы.

Флуктуационно-диссипативная теорема реализует функциональную связь между функцией отклика и автокорреляционной функцией. В системах с медленной динамикой флуктуационно-диссипативная теорема нарушается и в качестве характеристики этого нарушения вводится флуктуационно-диссипативное отношение (ФДО)

$$X(t, t_w) = \frac{TR(t, t_w)}{\partial C(t, t_w) / \partial t_w}. \quad (2)$$

В состоянии равновесия $X(t > t_w \gg t_{rel}) = 1$. На неравновесном этапе старения $X(t, t_w) \neq 1$. Асимптотическое значение величины флуктуационно-диссипативного отношения $X(t, t_w)$

$$X^\infty = \lim_{t_w \rightarrow \infty} \lim_{t \rightarrow \infty} X(t, t_w) \quad (3)$$

является новой универсальной характеристикой неравновесной критической релаксации [Calabrese, Gambassi, 2005].

В данной работе исследуется влияние дефектов структуры на эффекты старения при неравновесной критической эволюции XY-модели из начального низкотемпературного состояния, когда основной вклад в неравновесную динамику системы вносят спиновые волны. Эффекты старения при эволюции структурно неупорядоченной системы из начального высокотемпературного состояния были исследованы ранее в работе [Прудников и др., 2015б]. Показано, что наличие дефектов структуры существенно замедляет неравновесную динамику системы по сравнению с поведением «чистой» системы [Abriet, Karevski, 2004; Berthier et al., 2001].

Гамильтониан структурно неупорядоченной двумерной XY-модели имеет следующий вид:

$$H = -\frac{J}{2} \sum_{\langle i,j \rangle} p_i p_j \mathbf{S}_i \mathbf{S}_j, \quad (4)$$

где \mathbf{S}_i — классический планарный спин; p_i — случайные числа заполнения; J — обменный интеграл. Суммирование ведётся по всем парам ближайших соседей. Дефекты распределяются по решётке равномерно с вероятностью $c_{imp} = 1 - p$, где p — спиновая концентрация. В работе были рассмотрены следующие спиновые концентрации: $p = 1, 0$ (чистая беспримесная система), $0,9$ и $0,8$. Значения температур перехода Березинского-Костерлица-Таулесса для данных систем $T_{BKT}(p)$: $T_{BKT}(p = 1, 0) = 0,893(2)$, $T_{BKT}(p = 0, 9) = 0,679(7)$, $T_{BKT}(p = 0, 8) = 0,485(4)$, были определены в работе [Prudnikov, Popov, 2014]. Температура задается при этом в единицах обменного интеграла J . Поведение системы исследовалось во всей низкотемпературной фазе Березинского с $T \leq T_{BKT}$.

Моделирование проводилось на квадратной решётке с линейным размером $L = 256$. В работе [Surungan, Okabe, 2005] рассматривались решётки с линейными размерами $L < 256$, и из представленных в ней размерных асимптотик для ряда равновесных характеристик можно сделать вывод, что для решётки с $L = 256$ выход измеряемых величин должен осуществляться уже на их асимптотические значения. В работе [Gray et al., 2000] исследовалось нарушение динамического скейлинга и динамическое поведение корреляционной длины системы. При этом результаты моделирования неравновесного поведения для решёток с $L < 256$ позволяют заключить, что при изучении релаксационных процессов системы на решётке с $L = 256$ эффекты размерного изменения величин можно не учитывать.

В работе [Pereira et al., 2003] исследовалось влияние одиночного примесного узла на процесс пиннинга вихревого возбуждения. Полученные результаты позволяют сделать вывод, что при $L > 200$ наблюдается уже достаточно слабая зависимость характеристик вихревой подсистемы от размера L . В работе (Berche et al., 2003) моделирование структурно неупорядоченной системы было ограничено размером $L = 128$, при этом наблюдалось хорошее соответствие полученных результатов с результатами аналитических расчётов.

Моделирование осуществлялось нами с использованием алгоритма Метрополиса, реализующего диссипативную динамику в рамках модели A динамического критического поведения по классификации Гальперина-Хоэнберга [Folk, Moser, 2006; Hohenberg, Halperin, 1977]. Динамика Метрополиса адекватно реализует динамическую релаксацию в двумерной XY-модели во всей низкотемпературной фазе вплоть до точки фазового перехода $T \leq T_{BKT}$ (Прудников и др., 2014). Время в процессе моделирования измерялось в шагах Монте-Карло на спин (MCS/s), при этом шаг MCS/s определялся как временной промежуток, за который каждому спину в системе даётся возможность пробного изменения конфигурации.

Функция отклика (1) не может быть определена непосредственно методами компьютерного моделирования, поэтому более удобной диссипативной величиной является интегральная характеристика — динамическая обобщённая восприимчивость системы $\chi(t, t_w)$, для вычисления которой был применён метод воздействия на систему малых случайных бимодальных полей [Berthier et al., 2001].

В работе [Прудников и др., 2015б] было показано, что для корректного исследования асимптотики предела $t \rightarrow \infty$ и $t_w \rightarrow \infty$ необходим расчёт большого количества значений времени ожидания t_w , в противном случае возможен выход на неуниверсальную зависимость и получение недостоверного результата. Первоначально, для исследования старения системы из начального низкотемпературного состояния нами было использовано 16 значений времени ожидания: $t_w = 10, 20, 30, 40, 50, 60, 80, 100, 200, 300, 400, 500, 800, 1000, 2000$ и 5000 MCS/s при величине времени наблюдения $t - t_w = 15\,000$ MCS/s. Однако в последующем выяснилось, что этих значений недостаточно и для осуществления корректной процедуры расчёта флуктуационно-диссипативного отношения был рассмотрен беспрецедентный набор больших значений $t_w = 10\,000, 15\,000, 20\,000, 30\,000, 40\,000$ и $50\,000$ MCS/s при величине $t - t_w = 200\,000$ MCS/s. Для случая чистой системы усреднение проводилось по 12 000 различным реализациям динамики системы. Для неупорядоченной системы усреднение проводилось по 10 000 конфигурациям дефектов структуры и по 15 статистическим прогонкам для каждой примесной конфигурации.

На рис. 1 представлены результаты моделирования эффектов старения на примере двухвременной зависимости автокорреляционной функции $C(t, t_w)$ от времени наблюдения $t - t_w$ и от времени ожидания t_w . На вставке рис. 1а представлен график динамической скейлинговой зависимости

$$C(t, t_w) = t_w^{\eta/2} \Phi[\xi(t - t_w)/\xi(t_w)], \quad (5)$$

где η — величина аномальной размерности системы (критический индекс Фишера), значения которого для неупорядоченной были рассчитаны в работе [Прудников и др., 2015б]; $\xi(t)$ — временная зависимость корреляционной длины. В случае эволюции из начального низкотемпературного состояния $\xi(t) \approx \sqrt{t}$ [Berthier et al., 2001; Bray et al., 2000]. Из результатов, представленных на вставке к рис. 1а, можно заключить, что на долговременном этапе с $t - t_w \gg t_w$ динамическая скейлинговая зависимость (5) нарушается, и данные с разными t_w не ложатся на универсальную кривую.

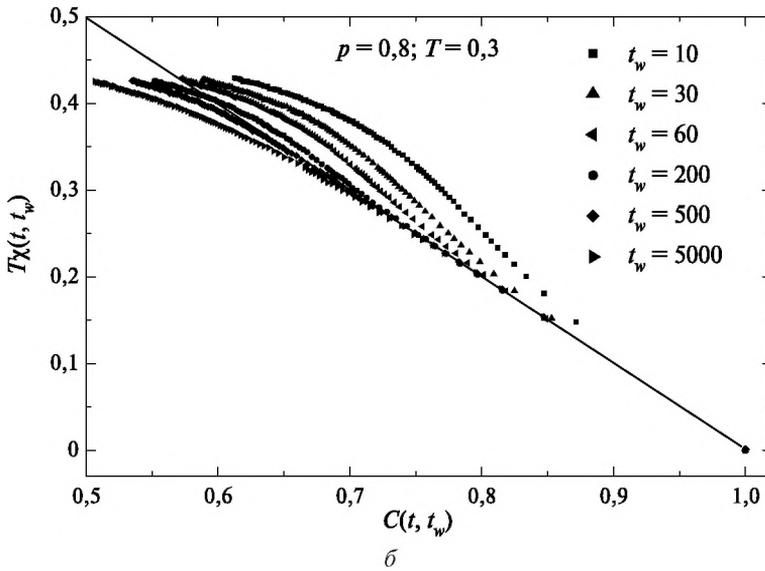
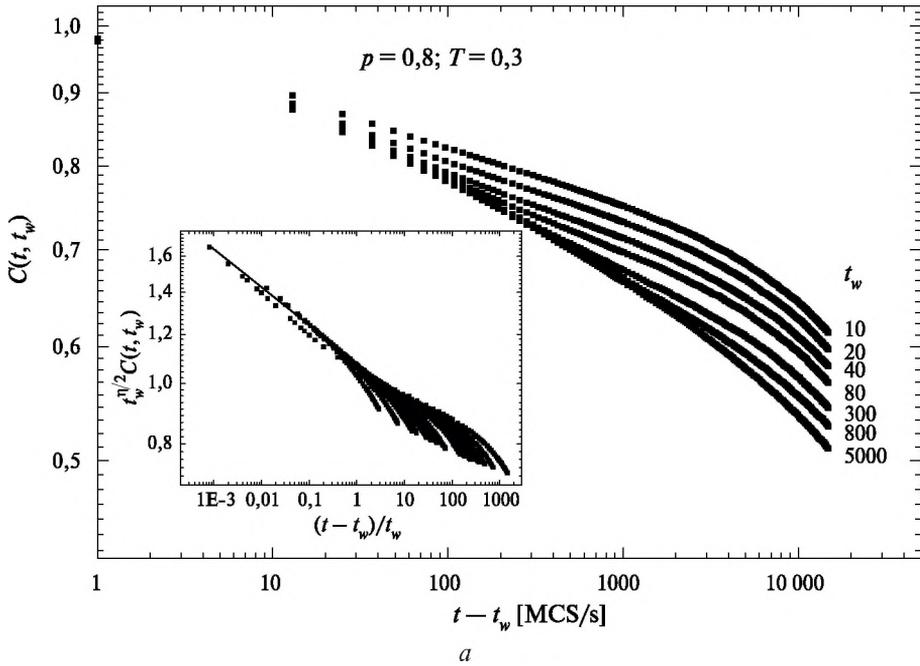


Рис. 1. Двухвременная зависимость автокорреляционной функции $C(t, t_w)$ для температуры $T=0,3$ и спиновой концентрации $p = 0,8$. На вставке представлены зависимости $t_w^{n/2} C(t, t_w)$ от $(t - t_w)/t_w$ для демонстрации скейлингового поведения автокорреляционной функции $C(t, t_w)$ и отсутствия «коллапса» данных на долговременном этапе (а). Параметрическая зависимость восприимчивости $T\chi(t, t_w)$ от $C(t, t_w)$. Прямая линия демонстрирует выполнение флуктуационно-диссипативной теоремы (б)

Параметрическая зависимость восприимчивости $T\chi(t, t_w)$ от $C(t, t_w)$ представлена на рис. 1б. Из приведённых данных можно сделать вывод, что для малых значений времени ожидания t_w величина $X(t_w) > 1$. Это согласуется с результатами работы [Berthier et al., 2001]. Однако нами выявлено, что с увеличением времени ожидания $X(t_w) \leq 1$.

Реализация расчёта предельного значения величины ФДО $X^\infty = \lim_{t_w \rightarrow \infty} X(t_w)$ показана на рис. 2 для различных значений времени ожидания t_w , температуры T и спиновой концентрации p . На основе представленных результатов (см. вставку на рис. 2) можно сделать вывод об отсутствии корректной асимптотики $X^\infty(t_w \rightarrow \infty)$.

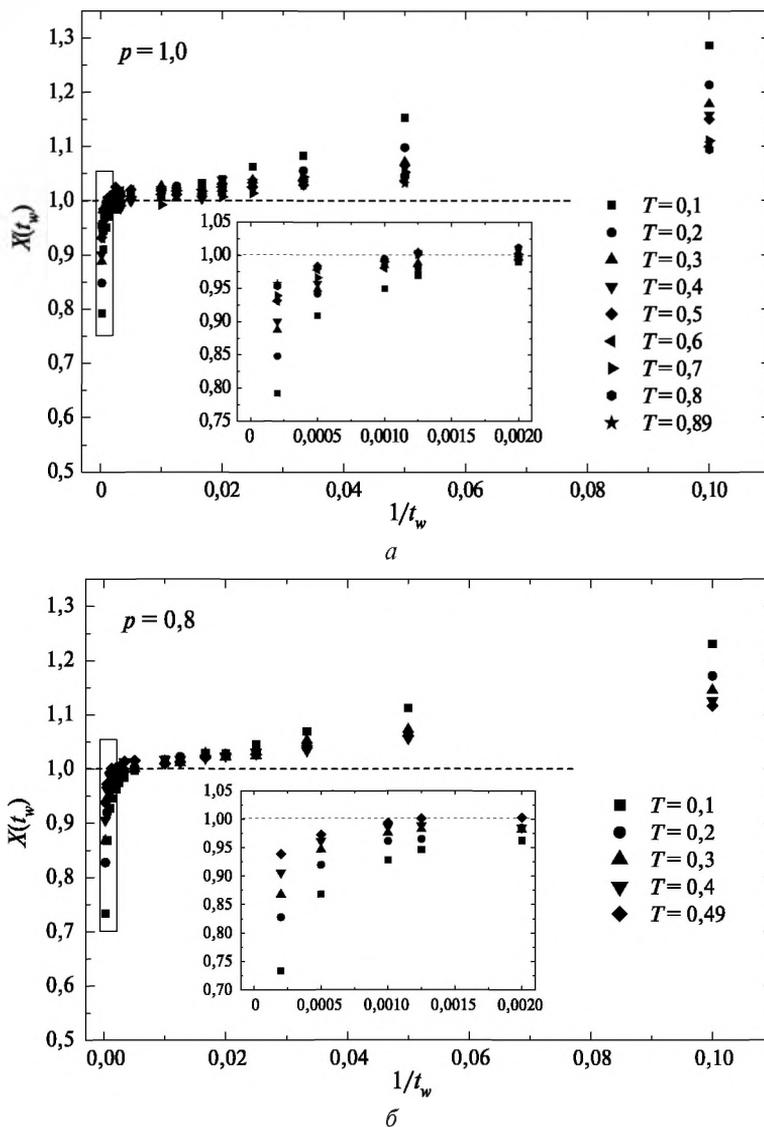


Рис. 2. Зависимость величины флуктуационно-диссипативного отношения $X(t_w)$ от $1/t_w$. Демонстрируется отсутствие асимптотики при реализации предела $t_w \rightarrow \infty$

Для выяснения возникшей ситуации был осуществлён расчёт динамических характеристик для больших значений $t_w \gg 10\,000$ MCS/s при $t - t_w = 200\,000$ MCS/s. Результаты моделирования представлены на рис. 3. Временные зависимости (вставка на рис. 3а) указывают на возможность осуществления в неравновесной релаксации системы эффектов субстарения (subaging) или суперстарения (superaging) [Henkel, Pleimling, 2010].

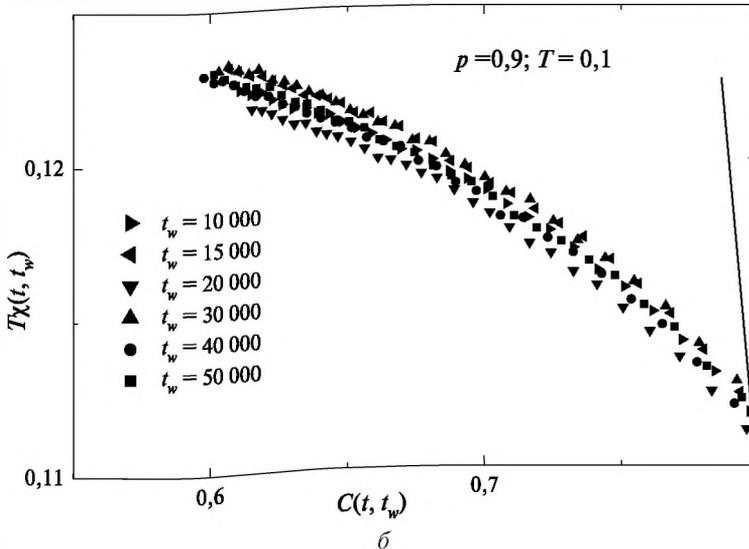
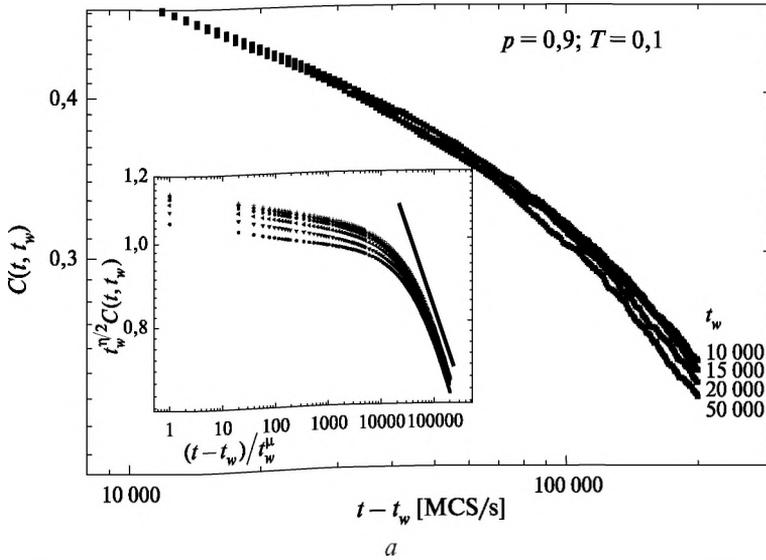


Рис. 3. Двухвременная зависимость автокорреляционной функции $C(t, t_w)$ для температуры $T = 0,3$ и спиновой концентрации $p = 0,8$ на больших значениях времени ожидания t_w . На вставке представлен динамический скейлинговый анализ эффектов субстарения (а). Параметрическая зависимость $T\chi(t, t_w)$ от $C(t, t_w)$ (б)

Динамическая скейлинговая зависимость для автокорреляционной функции в этом случае будет иметь вид

$$C(t, t_w) = t_w^{\eta(T, p)/2} \Phi \left[\xi(t - t_w) / \xi^{\mu(T, p)}(t_w) \right], \quad (6)$$

где μ — индекс субстарения ($\mu < 1$) или суперстарения ($\mu > 1$). При $\mu = 1$ осуществляется каноническое старение.

Возможность реализации данной скейлинговой зависимости была нами исследована, и удалось добиться «коллапса» данных для $C(t, t_w)$ с $\mu < 1$ (вставка на рис. 3а). Это говорит об осуществлении эффекта субстарения в процессе релаксации системы из низкотемпературного начального состояния. Следует отметить, что динамическая скейлинговая форма для субстарения реализуется на достаточно долговременном этапе наблюдения с $t - t_w \approx 70\,000 \dots 200\,000$ MCS/s. В пределах погрешности аппроксимации было установлено соотношение $\mu(T, p) = \eta(T, p)$ для всех рассмотренных T и p , которое указывает на некоторую внутреннюю связь аномальной размерности системы, как свойства, определяющего возможность осуществления фазового перехода в двумерной системе, и субстарения системы (т. е. эффекта неравновесной критической релаксации системы).

На временных участках, где реализуется динамический скейлинг с эффектом субстарения, был выполнен расчёт $X^\infty = \lim_{t_w \rightarrow \infty} X(t_w)$, процедура реализации кото-

рого представлена на рис. 4. Из приведённых результатов видно, что для них наблюдается корректная асимптотика $t_w \rightarrow \infty$. С использованием данной экстраполяции осуществлено вычисление предельного значения ФДО $X^\infty(T, p)$. Из представленных на рис. 5 результатов видно, что зависимость $X^\infty(T, p)$ линейна по температуре T . Значения $X^\infty(T_{BKT}, p)$ в точке фазового перехода Березинского-Костерлица-Таулесса $T_{BKT}(p)$ в пределах погрешности вычислений совпадают и равны $X^\infty(T_{BKT}, p) = 0,73(5)$ для всех рассмотренных спиновых концентраций p . Выявлена универсальная зависимость $X^\infty(T, p)$ как функция приведённой температуры $T/T_{BKT}(p)$, которая носит линейный характер для всех рассмотренных спиновых концентраций p . Данная температурная зависимость представлена на вставке к рис. 5.

Полученный результат позволяет сделать вывод, что $X^\infty(T, p)$ для всех рассмотренных спиновых концентраций p переопределяется только через температуру фазового перехода $T_{BKT}(p)$. При этом для одинаковых приведённых температур $T/T_{BKT}(p)$ значения $X^\infty(T, p)$ совпадают.

Выполненные исследования показывают, что релаксация двумерной XY-модели из начального низкотемпературного состояния существенно отличается от релаксации из начального высокотемпературного состояния [Прудников и др., 2015б]. Существование универсальной зависимости предельного ФДО от приведённой температуры означает, что основной вклад в неравновесную критическую релаксацию однородной и структурно неупорядоченной двумерной XY-модели при эволюции из начального низкотемпературного состояния вносят ангармонические спин-волновые эффекты, а вкладом вихревой подсистемы можно пренебречь.

В данной работе впервые осуществлён расчёт температурной зависимости асимптотического значения флуктуационно-диссипативного отношения $X^\infty(T, p)$ для однородной и структурно неупорядоченной двумерной XY-модели. Рассмотрен широкий диапазон спиновой концентрации p . Температурные зависимости получены для всей низкотемпературной фазы вплоть до точки фазового перехода Березинского-Костерлица-Таулесса $T \leq T_{BKT}$.

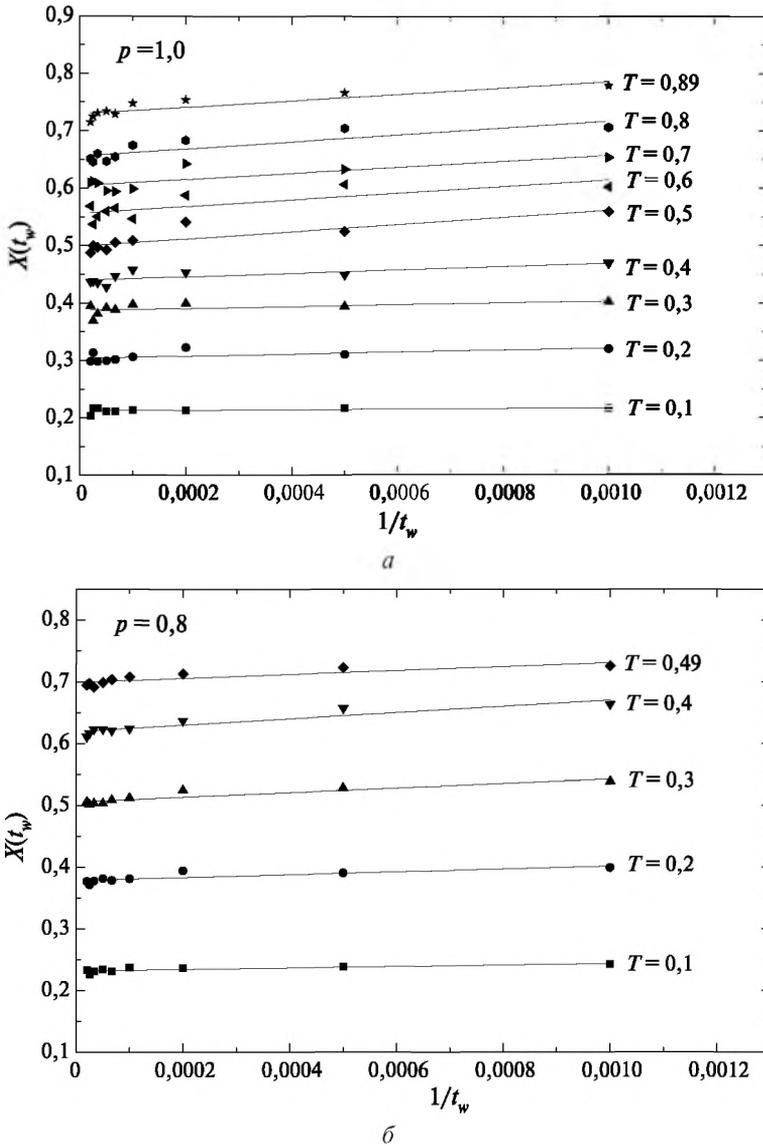


Рис. 4. Определение $X^\infty(T, p)$ путём корректной экстраполяции $1/t_w \rightarrow 0$ величин $X(t_w)$ для различных спиновых концентраций $p = 1,0$ (а) и $p = 0,8$ (б)

Впервые была получена универсальная зависимость $X^\infty(T, p)$ от приведённой температуры $T/T_{BKT}(p)$ для различных спиновых концентраций p , которая позволила установить, что основной вклад в процесс релаксации системы из низкотемпературного начального состояния вносят спиновые волны. Сравнение $X^\infty(T, p)$ в точке фазового перехода с аналитическим результатом позволяет сделать вывод, что основную роль играют ангармонические спин-волновые эффекты. Впервые было показано, что скейлинговая форма для двухвременных характеристик системы при релаксации из начального низкотемпературного состояния соответствует явлению субстарения.

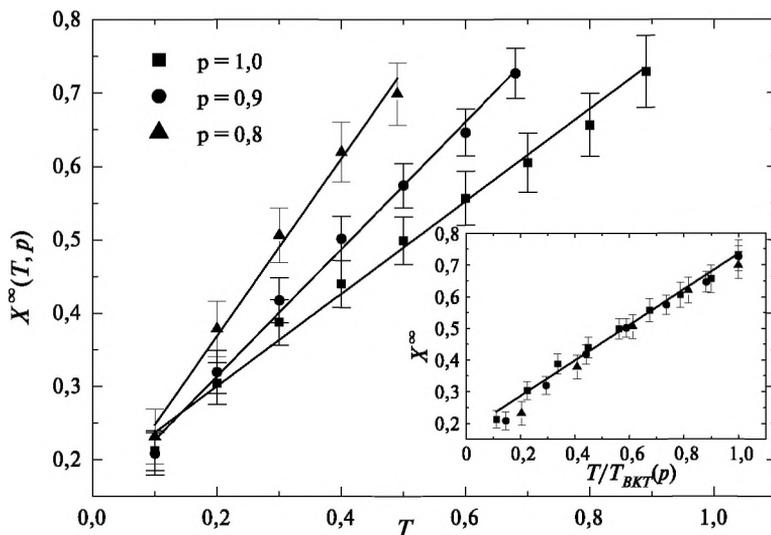


Рис. 5. Температурная зависимость асимптотического значения величины флуктуационно-диссипативного отношения $X^\infty(T, p)$ для различных спиновых концентраций p . На вставке — универсальная зависимость $X^\infty(T, p) \approx f[T/T_{\text{BKT}}(p)]$

ЛИТЕРАТУРА

- [Березинский, 1970] *Березинский В. Л.* Разрушение дальнего порядка в одномерных и двумерных системах с непрерывной группой симметрии // *ЖЭТФ*. 1970. Т. 59. С. 907–920.
- [Березинский, 2007] *Березинский В. Л.* Низкотемпературные свойства двумерных систем с непрерывной группой симметрии. М.: Физматлит, 2007. 232 с.
- [Коршунов, 2006] *Коршунов С. Е.* Фазовые переходы в двумерных системах с непрерывным вырождением // *УФН*. 2006. Т. 176. Вып. 3. С. 233–274.
- [Прудников и др., 2014] *Прудников В. В., Прудников П. В., Алексеев С. В., Попов И. С.* Исследование эффектов старения и температурной зависимости поперечной жёсткости в двумерной XY-модели // *ФММ*. 2014. Т. 115. Вып. 12. С. 1254–1261.
- [Прудников и др., 2015a] *Прудников В. В., Прудников П. В., Вакилов А. Н., Попов И. С.* Теоретические методы описания неравновесного критического поведения структурно неупорядоченных систем и эффектов старения. Омск: Изд-во ОмГУ, 2015. 334 с.
- [Прудников и др., 2015b] *Прудников В. В., Прудников П. В., Попов И. С.* Неравновесные эффекты старения в критическом поведении структурно неупорядоченных планарных магнетиков // *Письма в ЖЭТФ*. 2015. Т. 101. Вып. 8. С. 596–601.
- [Abriet, Karevski, 2004] *Abriet S., Karevski D.* Off equilibrium dynamics in the 2d-XY system // *Eur. Phys. J. B*. 2004. V. 37. P. 47–53.
- [Berche et al., 2003] *Berche B., Farinas-Sanchez A. I., Holovatch Yu., Paredes R. V.* Influence of quenched dilution on the quasi-long-range ordered phase of the 2d XY model // *Eur. Phys. J. B*. 2003. V. 36. P. 91–98.
- [Berthier, Kurchan, 2013] *Berthier L., Kurchan J.* Non-equilibrium glass transitions in driven and active matter // *Nature Phys.* 2013. V. 9. P. 310–314.
- [Berthier et al., 2001] *Berthier L., Holdsworth P. C. W., Sellitto M.* Nonequilibrium critical dynamics of the two-dimensional XY model // *J. Phys. A: Math. Gen.* 2001. V. 34. P. 1805–1824.

- [Bray et al., 2000] *Bray A. J., Briant A. J., Jervis D. K.* Breakdown of Scaling in the Nonequilibrium Critical Dynamics of the Two-Dimensional XY Model // *Phys. Rev. Lett.* 2000. V. 84. P. 1503–1506.
- [Calabrese, Gambassi, 2005] *Calabrese P., Gambassi A.* Ageing properties of critical systems // *J. Phys. A.* 2005. V. 38. P. R133–R193.
- [Folk, Moser, 2006] *Folk R., Moser G.* Critical dynamics: a field-theoretical approach // *J. Phys. A: Math. Gen.* 2006. V. 39. P. R207–R313.
- [Henkel, Pleimling, 2010] *Henkel M., Pleimling M.* Non-Equilibrium Phase Transitions, Vol. 2: Ageing and Dynamical Scaling far from Equilibrium, Theoretical and Mathematical Physics. Springer: Heidelberg, 2010. 544 p.
- [Hohenberg, Halperin, 1977] *Hohenberg P. C., Halperin B. I.* Theory of dynamics critical phenomena // *Rev. Mod. Phys.* 1977. V. 49. P. 435–479.
- [Kosterlitz, 1974] *Kosterlitz J. M.* The critical properties of the two-dimensional XY model // *J. Phys. C.: Solid State Phys.* 1974. V. 7. P. 1046–1060.
- [Kosterlitz, Thouless, 1973] *Kosterlitz L. M., Thouless D. J.* Ordering, metastability and phase transitions in two-dimensional systems // *J. Phys. C.* 1973. V. 6. P. 1181–1203.
- [Surungan, Okabe, 2005] *Surungan T., Okabe Y.* Kosterlitz-Thouless transition in planar spin models with bond dilution // *Phys. Rev. B.* 2005. V. 71. P. 184438.
- [Pereira et al., 2003] *Pereira A. R., Mol L. A. S., Leonel S. A., Coura P. Z., Costa B. V.* Vortex behavior near a spin vacancy in two-dimensional XY magnets // *Phys. Rev. B.* 2003. V. 68. P. 132409.
- [Prudnikov, Popov, 2014] *Prudnikov P. V., Popov I. S.* Non-equilibrium critical dynamics in pure and diluted 2D XY-model // *J. Phys.: Conf. Series.* 2014. V. 510. P. 012014.
- [Prudnikov, Popov, 2015] *Prudnikov P. V., Popov I. S.* Coarsening in Critical Dynamics of 2D XY model // *Solid State Phenomena.* 2015. V. 233–234. P. 8–11.
- [Surungan, Okabe, 2005] *Surungan T., Okabe Y.* Kosterlitz-Thouless transition in planar spin models with bond dilution // *Phys. Rev. B.* 2005. V. 71. P. 184438.
- [Tauber, 2014] *Tauber U. C.* Critical Dynamics: A Field Theory Approach to Equilibrium and Non-Equilibrium Scaling Behavior. Cambridge: Cambridge University Press, 2014. 529 p.

MONTE-CARLO SIMULATION OF NON-EQUILIBRIUM CRITICAL BEHAVIOR OF TWO-DIMENSIONAL DILUTED XY-MODEL

I. S. Popov, P. V. Prudnikov, V. V. Prudnikov

Omsk State University (OSU)

Numerical simulation of non-equilibrium critical relaxation of diluted two-dimensional XY-model was performed in the low-temperature Berezinskii phase for wide range of spin concentrations. Simulation was carried out for relaxation from low-temperature non-equilibrium initial states by Monte-Carlo methods. Subaging effects and violation of the fluctuation-dissipation theorem were revealed.

Keywords: two-dimensional XY-model, non-equilibrium critical behavior, Berezinskii-Kosterlitz-Thouless phase transition, structure disorder, aging effects, fluctuation-dissipation ratio, Monte-Carlo methods.

Popov Ivan Sergejevich — postgraduate student, diphosgen@mail.ru

Prudnikov Pavel Vladimirovich — professor, doctor of physical and mathematical sciences, prudnikp@univer.omsk.su

Prudnikov Vladimir Vasilevich — head of department, doctor of physical and mathematical sciences, prudnikov@univer.omsk.su

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ФРАГМЕНТАЦИИ И АБЛЯЦИИ МОЛЕКУЛЯРНОГО ОБЛАКА СИЛЬНОЙ УДАРНОЙ ВОЛНОЙ

Б. П. Рыбакин

Федеральный научный центр Научно-исследовательский институт системных исследований
Российской академии наук (ФНЦ НИИСИ РАН)

Московский государственный университет имени М. В. Ломоносова (МГУ им. Ломоносова)

В работе представлены результаты компьютерного моделирования взаимодействия ударной волны с молекулярным облаком в трёхмерной постановке. Сжатие облаков, вызванное ударными волнами, джетами и другими процессами часто встречается в Галактиках. В результате такого взаимодействия молекулярные облака (МО) либо разрушаются, либо в них ускоряются процессы образования звёзд и планетных систем. Ударные волны, которые распространяются в межзвёздной среде (МЗС), образуются в результате различных астрофизических процессов, например, джетов (jets), взрывов сверхновых звёзд, столкновений МО и т. д. Для проведения моделирования была построена трёхмерная параллельная программа с достаточным эффективным разрешением для диаметра облака. Проведён детальный анализ процессов абляции, деформирования и фрагментации молекулярного облака. Подробно изучено формирование системы прошедших и отражённых волн, сжатие и абляция вещества. Постпроцессорная обработка результатов расчётов позволила найти особенности возникающего течения — образование вихрей, эрозии и абляции вещества молекулярного облака.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 14-29-06055).

Ключевые слова: математическое моделирование, ударные волны, молекулярные облака, параллельные вычисления, OpenMP.

ВВЕДЕНИЕ

Пространство между звёздами в галактиках заполняет разреженное вещество, которое называется межзвёздной средой (МЗС). Оно состоит из газа и пыли, и составляет 10...15 % от общей массы в галактическом диске [Ferriere, 2001]. Из него образуются звёзды, которые продолжают формировать МЗС на протяжении своего жизненного цикла и после гибели. Межзвёздная среда по своей структуре существенно неоднородна. В МЗС существуют области, в которых вещество имеет большую плотность. Эти неоднородности возникают по разным причинам, но независимо от природы возникновения, такие сгустки или облака (МО) качественно отличаются от МЗС. Сжатие молекулярных облаков (МО) — один из важных механизмов, ускоряющих процесс гравитационного коллапса и формирование звёзд и планетных систем. Так же оно может привести к разрушению МО указанными воздействиями. На рис. 1 приведён снимок молекулярного облака Unicorn R2, полученного с помощью телескопа VISTA ESO.

Исследованию особенностей поведения молекулярно-пылевых облаков и изучению их роли в формировании звёзд посвящено большое количество работ. В пионерских работах [Hartquist, Dyson, 1988; Hartquist et al., 1997] аналитиче-

ски были проанализированы процессы взаимодействия потоков с различными неоднородностями, изучены вопросы, связанные с влиянием внешних течений на свойства неоднородных областей. Первые работы в этой области выполнены в одномерной или двумерной осесимметричной постановках. По мере роста вычислительных мощностей появились исследования, в которых добавились учёт гравитации и влияние магнитных полей. В работах [Anderson et al., 1994; Gregori et al., 1999; Klein et al., 1994; Truelove et al., 1998] была исследована задача взаимодействия потока вещества и ударной волны с одиночным МО. Эти исследования также проводились в одномерной и двумерной постановках. Было исследовано влияние основных гидродинамических характеристик и магнитных полей. Одномерным и двумерным постановкам характерны ограничения, которые не позволяют провести детальное исследование процесса взаимодействия ударной волны с МО.

В трёхмерной постановке такие ограничения отсутствуют. Это даёт дополнительные преимущества, связанные с более точным описанием возникающих процессов. В работах [Frank et al., 2003; Melioli et al., 2005; Stone, Norman, 1992; Truelove et al., 1997] была использована двух и трёхмерная постановка задачи. В этих и ряде других исследований появилась возможность выявить сложные механизмы, возникающие в процессе взаимодействия сильных ударных волн с молекулярно-пылевыми облаками. Изучено появление рэлей-тейлоровской-неустойчивости, явление уноса вещества (абляции) при образовании турбулентного течения и радиационного охлаждения вещества облака. В работе [Frank et al., 2003] используется однофазная модель газа, которая достаточно точно отражает физическую картину течения. В расчётах предполагалось, что невозмущённая субстанция МЗС состоит из сравнительно тёплого вещества ($\sim 10^4$ К) и небольших неоднородно распределённых холодных облаков ($\sim 10^2$ К), которые обладают большой плотностью [Anderson et al. 1994; McKee, 1988; Melioli et al., 2005]. Изначально облака находятся в динамическом равновесии с фоновым газом [Frank et al., 2003].



Рис. 1. В центре молекулярное облако Unicorn R2. Telescope VISTA.
The authors and rights European Southern Observatory

В данной работе рассмотрен процесс взаимодействия ударной волны большой интенсивности с молекулярным облаком. Моделируется процесс возникающего гиперзвукового газодинамического течения и анализируется взаимодействие процессов сжатия, фрагментации и образование турбулентных течений в облаке и окружающей среде.

Моделирование динамических процессов в трёхмерной постановке на достаточно подробных сетках требует больших вычислительных ресурсов. Для проведения расчётов был разработан параллельный алгоритм решения задач нестационарной газодинамики [Rybakin, 2013], детально опробованный при решении задачи взаимодействия набегающей ударной волны с областью меньшей плотности [Rybakin, Goryachev, 2014]. Для ускорения расчётов была использована технология OpenMP. Для этого была разработана параллельная программа, которая основана на разностных схемах высокого разрешения. Данная программа позволяет решать трёхмерные уравнения Эйлера для сжимаемого газа. Для улучшения качества параллельного кода использовался Intel V Tune Amplifier XE.

МЕТОДИКА И РЕЗУЛЬТАТЫ РАСЧЁТОВ

Межзвёздная среда состоит из сравнительно тёплого вещества $\sim 10^4$ К. В МЗС находится молекулярное облако, плотность намного больше плотности МЗС. Температура в молекулярном облаке меньше температуры МЗС $T_{cloud} \approx 10^2$ К. Ударная волна, образовавшаяся после взрыва сверхновой звезды, набегаёт на молекулярное облако, которое находится в межзвёздной среде. На рис. 2 схематически представлена геометрия расчётной области. Приведено сечение трёхмерной расчётной области плоскостью, перпендикулярной оси OZ . На рис. 2 приведено начальное ($t = 0$) положение ударной волны и МО, плоская ударная волна приближается к левой границе молекулярного облака (слева). На правом рисунке приведён момент безразмерного времени $t = 7$. В расчётах используется однофазная модель газа, которая обеспечивает детальную физическую картину.

Вычислительная область представляет собой параллелепипед с размерами от $2048 \times 1024 \times 1024$ до $1024 \times 512 \times 512$, размеры ячеек равны: $\Delta x = \Delta y = \Delta z$.

Ударная волна взаимодействует с МО, которое имеет другие термодинамическими характеристиками. Кроме того, ударная волна имеет конечную толщину, её поверхность не плоская, и уравнения состояния в межзвёздной среде, МО и ударной волне отличаются друг от друга. Каждый газ может иметь свою собственную скорость звука $a_i = \sqrt{\gamma_i(p_i/\rho_i)}$, молекулярную массу, показатель γ и т. д. Однако, как показано в работах [Bagabir, Drikakis, 2001; Picone, Boris, 1988], можно использовать одинаковые значения термодинамических характеристик для различных газов и использовать модель с $\gamma 5/3$.



Рис. 2. Цифровой шпирен для градиента плотности при $t = 0$ (слева) и при $t = 7$ (справа)

Будем рассматривать нестационарные течения идеального сжимаемого газа. Также будем пренебрегать эффектами вязкости и трения. Тогда система законов сохранения массы и количества движения и энергии в трёхмерной декартовой системе координат может быть записана следующим образом:

$$\frac{\partial U}{\partial t} + \frac{\partial F(U)}{\partial x} + \frac{\partial G(U)}{\partial y} + \frac{\partial H(U)}{\partial z} = 0. \quad (1)$$

Вектор консервативных переменных записывается следующим образом: $U = (\rho, \rho u, \rho v, \rho w, \rho e)^T$. Вектор потоков:

$$F(U) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ \rho uw \\ \rho ue + up \end{pmatrix}. \quad (2)$$

Потоки $G(U)$ и $H(U)$ можно записать аналогично (2).

Для замыкания системы уравнений используется уравнение состояния идеального газа:

$$p = \rho(\gamma - 1) \left| e - \frac{1}{2}(u^2 + v^2 + w^2) \right|, \quad (3)$$

здесь γ — отношение теплоёмкостей; ρ — плотность; $v = \{u, v, w\}$ — вектор скорости; p — давление [Rybakin, Goryachev, 2014].

Изначально облако находится в динамическом равновесии с фоновым газом. Теплопроводностью и радиационными потерями при взаимодействии можно пренебречь. В работе используется уравнение состояния идеального газа с $\gamma = 5/3$. Плотность межзвёздной среды принимается равной $\rho_a = 2,15 \cdot 10^{-25}$ г/см³, температура $T_a = 10^4$ К, $u_a = 0,0$. Плотность в облаке $\rho_c = 1,075 \cdot 10^{-22}$ г/см³, температура $T_c = 100$ К, $u_c = 0,0$. Параметры газа за ударной волной определяются из уравнений Ренкина-Гююнио.

В наших расчётах число Маха ударной волны равно $M = 7$, плотность $\rho_{sh} = 8,6 \cdot 10^{-22}$ г/см³, температура $T_{sh} = 1,5 \cdot 10^5$ К, скорость $u_{sh} = 1,04 \cdot 10^7$ см/с [Klein et al., 1994; McKee, 1988; Melioli et al., 2005].

Толщина фронта ударной волны достаточно большая 2...5 пк, что намного больше радиуса облака. Радиус облака равен $r_c = 0,05$ пк, время прохождения ударной волной расстояния диаметра облака приблизительно равно 960 лет.

Одним из наиболее важных параметров процесса взаимодействия ударной волны с МО является отношение плотности МО к плотности МЗС:

$$\chi = \frac{\rho_{mc}}{\rho_{ism}}, \quad (4)$$

здесь ρ_{MO} — плотность молекулярного облака; $\rho_{МЗС}$ — плотность межзвёздной среды соответственно. Предполагаем, что ударная волна не подвергается радиационному охлаждению и число Маха ударной волны достаточно большое.

Тогда можно показать, что скорость ударной волны в облаке можно приблизительно оценить [McKee, Cowie, 1975] так:

$$v_{c,sh} = \frac{v_{sh}}{\chi^{1/2}}. \quad (5)$$

Используя выражения (4) и (5) можно оценить время, за которое ударная волна достигнет центра облака:

$$\tau_{cc} = \frac{\chi^{1/22} R_c}{v_{sh}} \quad (6)$$

ОБЕЗРАЗМЕРИВАНИЕ

Запишем приведённую систему трёхмерных уравнений Эйлера в безразмерном виде. Для этого будем представлять размерные величины в виде $f = f_0 f'$. Здесь f' — размерная величина; f_0 — константа обезразмеривания. Все величины, представленные в работе, приведены к безразмерному виду.

НАЧАЛЬНЫЕ И ГРАНИЧНЫЕ УСЛОВИЯ

Приведённые в работе результаты были рассчитаны на сетке с размерами $1024 \times 512 \times 512$. Радиус молекулярного облака был равен 64 ячейкам, и его центр располагался в точке с координатами c_x, c_y, c_z . На левой и правой границах расчётной области использовались условия открытой границы. На остальных границах задавались периодические граничные условия. При заданных начальных условиях начинается расчёт процесса взаимодействия высокоинтенсивной ударной волны с молекулярным облаком. В начальный момент времени ударная волна примыкает к левой границе облака, и в момент времени $t = 0$ начинает взаимодействовать с облаком.

МЕТОД ЧИСЛЕННОГО РЕШЕНИЯ

Для решения задачи был создан параллельный алгоритм, использующий разностные схемы высокого разрешения, основанный на схемах типа ТВД [Годунов, 1959; Того, 2009]. Разностная схема имеет второй порядок точности и позволяет вести расчёт с высоким разрешением движения ударных волн и контактных разрывов и предотвращать нефизические осцилляции.

Для верификации программного кода было проведено большое количество различных двух и трёхмерных тестов [Rybakin, 2013, Rybakin, Goryachev, 2014]. Проведённое тестирование показало хорошее совпадение результатов расчётов с аналитическими решениями и экспериментальными данными. Фронты ударных волн размазывались на величину 3–4 ячейки, контактные разрывы — на 3–5 ячеек.

Для хорошего разрешения процессов, происходящих в МО необходимо, чтобы радиус облака был не менее 64 расчётных ячеек. Размер разностной сетки был выбран из этих соображений, основные расчёты проводились на сетке размерами $1024 \times 512 \times 512$. Меньшее количество ячеек отрицательно сказывается на разрешении происходящих процессов.

Радиус вещества МО — r_C ; безразмерные значения газа — $\rho'_c = 500$. Фронт ударной волны, которая движется слева направо, в начальный момент времени примыкает к левой границе МО. Число Маха ударной волны равно $M = 7$. Величины за фронтом ударной волны находятся из условий Rankine-Hugoniot с $\gamma = 5/3$.

Шаг по времени определяется из условий устойчивости Courant-Friedrichs-Lewy. Проведённая тщательная верификация программы показала хорошее совпадение с аналитическими решениями и экспериментальными данными [Rybakin, Goryachev, 2014; Rybakin et al., 2014].

РЕЗУЛЬТАТЫ

В процессе взаимодействия ударной волны с облаком выделяют несколько характерных стадий. На начальной стадии под действием падающей ударной волны облако сжимается. Разность величин в МЗС, молекулярном облаке и за фронтом ударной волны меняется на несколько порядков. Поэтому на приведённых графиках используется логарифм плотности, энергии и температуры.

На рис. 3 приведён численный шпирен для логарифма плотности на момент безразмерного времени $t = 50$. На рис. 3 отражены основные характеристики возникающего течения. Условные обозначения: M1 — первичная ударная волна (SW), отражённая от основной ударной волны, возникшая при отражении Маха на оси симметрии; M2 — вторичные SW; S1 — первичные ножки Маха; S2 — вторичные ножки Маха; T1 — первичные тройные точки; T2 — вторичные тройные точки; V1 — первичные отрывные волны, вызванные регулярным отражением от вогнутого фронта основной ударной волны; V2 — вторичные вихревые образования — волны, вызванные отражением от вогнутого фронта основной ударной волны и выпуклости в области образования ножек Маха. После прохождения ударной волной $L = 3/4$ диаметра облака начинается процесс его фрагментации, и активизируется совместный процесс абляции — перемешивание МО с областью ударной волны и межзвёздной средой. В левом нижнем углу приведена изоповерхность логарифма плотности в момент времени после прохождения ударной волной облака, цвет на поверхности отображает модуль местной скорости течения.

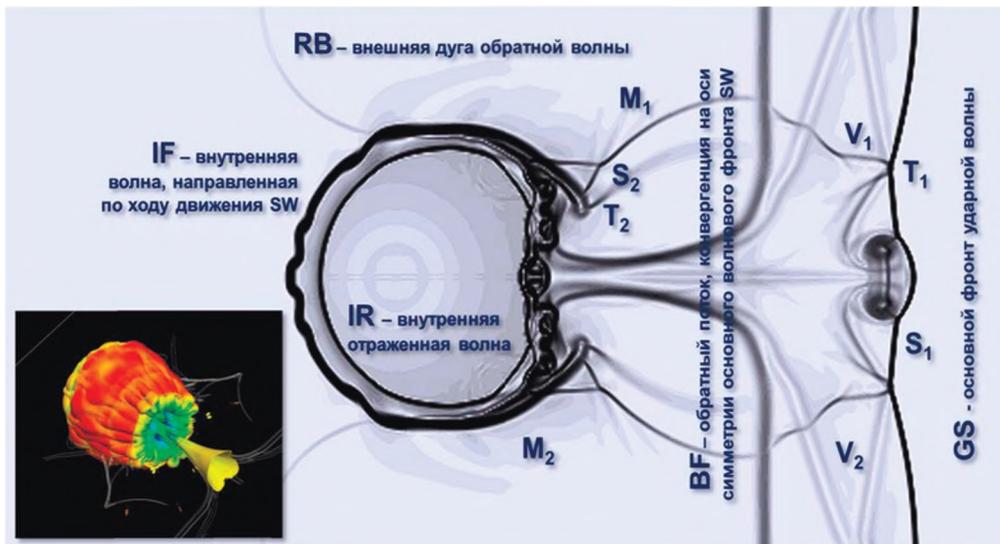


Рис. 3. Цифровой шпирен для градиента плотности при $t = 50$ и изображение одной из изоповерхностей модуля ротора векторного поля скорости, с приведённой на ней картой распределения локальной скорости потока газа

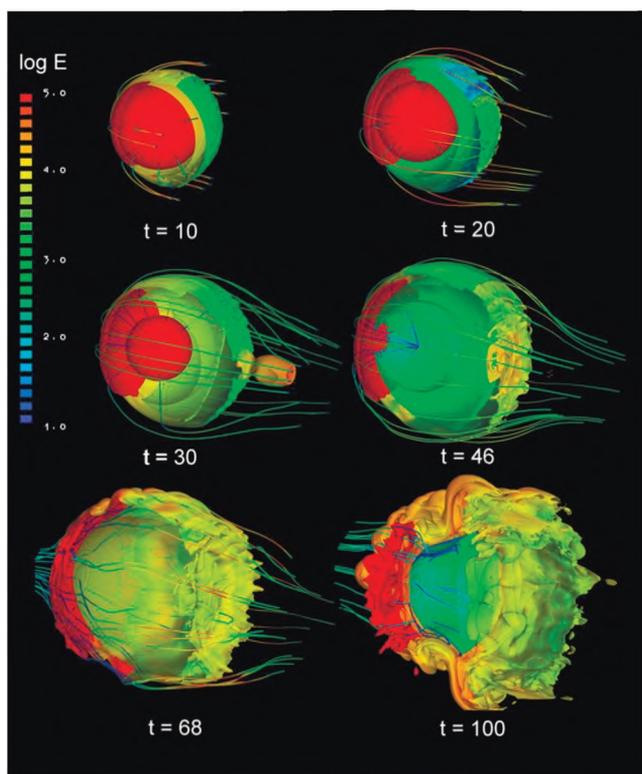


Рис. 4. Цифровой шпирен для градиента плотности при $t = 50$ и изображение одной из изоповерхностей модуля ротации скоростного поля, с приведённой на ней картой распределения локальной скорости потока газа

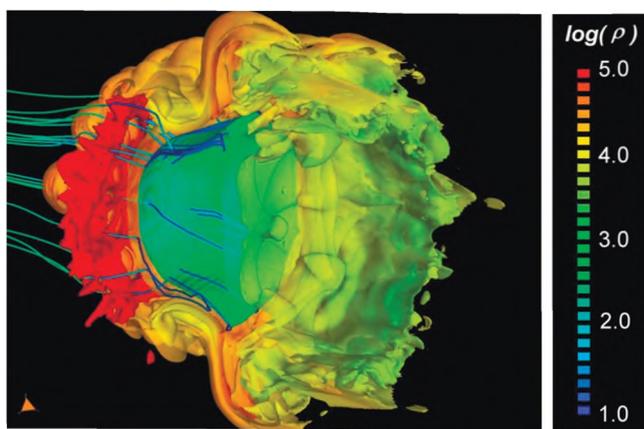


Рис. 5. Изоповерхность плотности для $\log r = 2, 4, 5$. На изоповерхностях цветом отображено распределение энергии. Стрелками показаны линии тока

В начальные моменты времени, когда головная ударная волна начинает огибать молекулярное облако, на левой границе образуется волна, которая движется навстречу потоку. За счёт этого на левой границе облака начинает формироваться первичное возмущение. На правой границе возникает неустойчивость Рихтмаейра-Мешкова (Р-М). Следующая стадия характеризуется расширением облака и образованием сложной волновой структуры. За фронтом прошедшей ударной волны формируются вихри, которые развиваются с течением времени. На поверхности облака виден процесс абляции (уноса вещества).

После прохождения ударной волной сферического облака профиль ударной волны (УВС) становится вогнутым, на передней кромке видны тройные ножки Маха (см. рис. 3). Падающая ударная волна распадается на проходящую и отражённую ударную волны. На поздней стадии после прохождения ударной волной всего облака на оси тыльной поверхности облака в области отражённой ударной волны формируется область высокого давления. Эта область высокого давления образует ударную волну, которая начинает распространяться навстречу падающей волне.

На рис. 4 приведены изоденсы для значения плотности $\log r = 2, 4, 5$. На этой поверхности цветом отображено значение энергии, приведены линии тока для выделенных частиц. Они изображаются линиями со стрелками. На левой стороне облака по оси X хорошо видна область высокого давления.

Прошедшая ударная волна «сдувает» возмущения с поверхности молекулярного облака. На рис. 5 можно отметить образование поверхностных возмущений. На рисунке представлено значение ротора поля скоростей. Цветом показано распределение локальной скорости потока газа на этой поверхности. Неустойчивость Рихтмаейра-Мешкова является частным случаем неустойчивости Рэля-Тейлора и возникает из-за нарастания возмущения на границе раздела двух сред с различными плотностями, подвергнутых воздействию ударной волны. На рис. 5 видно формирование неустойчивости Р-М на правой границе. На правой половине сферы, за фронтом ударной волны, видно формируемое вихревое кольцо, которое с течением времени развивается и образует вихревую структуру.

Важной особенностью процесса взаимодействия ударной волны с молекулярным облаком является образование вихрей. Ударно-волновое взаимодействие образует систему вихрей двумя способами:

- бароклинный эффект создаёт сдвиговой слой между облаком и окружающей средой;
- эффект искривлённой ударной волны создаёт турбулентность в диффузной среде в непосредственной близости от облака.

Возникающие возмущения приводят к образованию завихренности над поверхностью облака и уносу (абляции) его вещества. Это хорошо видно на рис. 6. На этом рисунке представлена визуализация изоповерхности для второго инварианта тензора скорости — Q -критерия [Hunt et al., 1988].

Вихри образуют кольцевые структуры на поверхности облака. На рис. 7 приведены результаты расчётов в «поздние» моменты времени. Зарождающиеся вихри превращаются в вихревые дорожки и инициируют вихревые шпильки. Соседние вихри с противоположным направлением вращения образуют систему первичных и вторичных вихревых структур.

Со временем эти вихревые структуры растягиваются, изгибаются и меняют форму, образуя вторичные, третичные и т. д. системы. На рис. 7 показан квази-продольный вихрь (quasi-streamwise vortex). Хорошо видны вихревые шпильки, красным цветом показан пучок удлинённых вихрей.

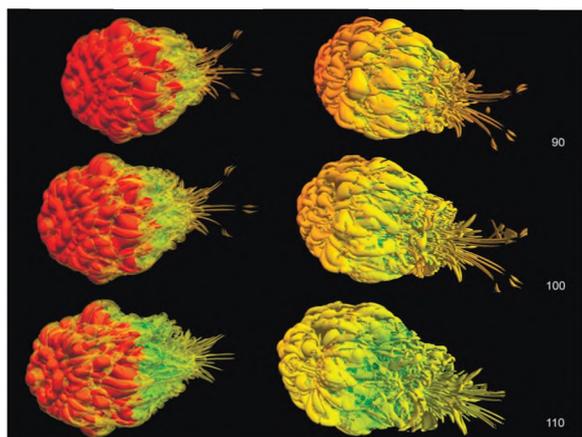


Рис. 6. Возмущения в приповерхностных слоях отражаются с использованием идентификации вихрей посредством Q-критерия



Рис. 7. Неустойчивость Рихтмайера-Мешкова — интенсивное вихреобразование — фрагментация по плотности газа облака после прохождения ударной волны

ПАРАЛЛЕЛЬНЫЙ АЛГОРИТМ

Математическое моделирование трёхмерных задач нестационарной газовой динамики на достаточно больших сетках требуют больших вычислительных ресурсов. Для сокращения времени расчетов был построен параллельный алгоритм и программа, которая позволяет проводить вычисления на гибридной вычислительной системе. Настройка производительности параллельного алгоритма проводилась с помощью Intel VTune Amplifier XE. Он позволяет выполнять профилирование приложений непосредственно на узле кластера. Был использован тип анализа Lightweight Hotspots. С помощью этого средства были определены функции, на выполнение которых тратится больше всего времени. Технология Lightweight Hotspots основана на использовании специальных аппаратных регистров (Performance Monitoring Units), которые предназначены для учёта различных низкоуровневых событий.

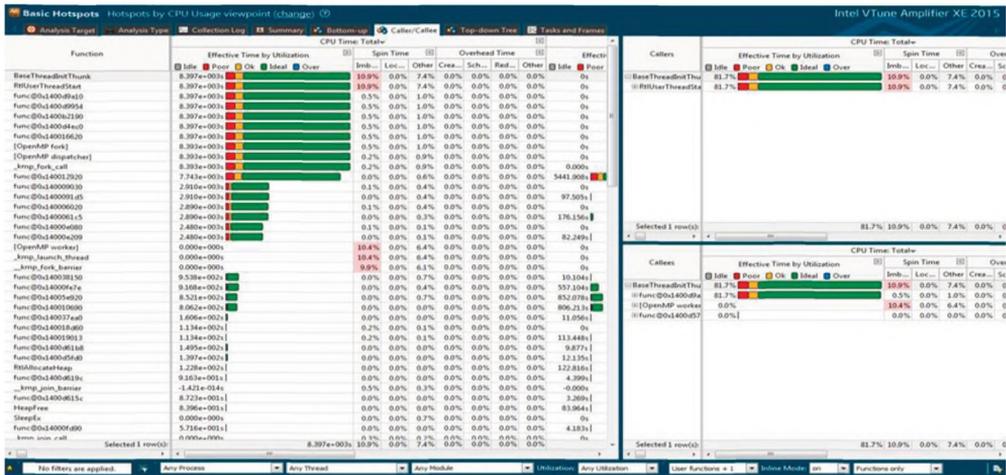


Рис. 8. Анализ качества распараллеливания программы, проведённый с помощью программы Amplifier XE

В процессоре Intel Xeon, которые были использованы в расчётах, на ядро приходится два регистра, накапливающих информацию о событиях, специфичных для потока или ядра. На рис. 8 приведены результаты проведённого анализа по распараллеливанию программы на CPU. На этом рисунке показано «качество» построенного параллельного алгоритма. Более 80 % общего времени занимает расчёт четырёх подпрограмм. После оптимизации этих подпрограмм их степень распараллеливания стала достаточно высокой.

На рис. 8 зелёным цветом показано «идеальное» распараллеливание, жёлтым — хорошее, красным — плохое. Плохое распараллеливание означает, что эффективность в этом случае меньше 50 %. Некоторые подпрограммы рассчитывались на графическом процессоре.

ВЫВОДЫ

Проведено численное моделирование процесса взаимодействия ударной волны с молекулярным облаком. Подробно изучен процесс абляции с поверхности облака и образование вихревых структур. В начальные моменты времени на левой границе облака начинает формироваться первичное возмущение. На правой границе возникает неустойчивость Рихтмайера-Мешкова. Следующая стадия характеризуется расширением облака и образованием сложной волновой структуры. За фронтом прошедшей ударной волны формируется вихри, которые развиваются с течением времени. На поверхности облака виден процесс абляции (уноса вещества). В моменты безразмерного времени, начиная с $t \approx 80$, образуется система вихрей. Вихри образуют кольцевые структуры на поверхности облака. С течением времени они превращаются в вихревые дорожки и инициируют вихревые шпильки. Соседние вихри с противоположным направлением вращения образуют систему первичных и вторичных вихревых структур. Со временем эти вихревые структуры растягиваются, изгибаются и меняют форму, образуя вторичные, третичные и т. д. структуры. Визуализация возникающих многомерных течений осуществлялась с помощью графического комплекса HDVIS [Rybakin, Goryachev, 2015].

ЛИТЕРАТУРА

- [Годунов, 1959] *Годунов С. К.* Разностный метод численного расчета разрывных решений уравнений гидродинамики // *Мат. сборник*. 1959. Т. 47(89). № 3. С. 271–300.
- [Anderson et al., 1994] *Anderson M. C., Jones T. W., Rudnick L., Tregillis I. L., Kang H.* The dynamical and radiative evolution of clumpy supernova ejecta // *Astrophysical J.* 1994. V. 421. No. 1. P. L31–L34.
- [Bagabir, Drikakis, 2001] *Bagabir A., Drikakis D.* Mach number effects on shock-bubble interaction // *Shock Waves*. 2001. V. 11. P. 209–218.
- [Ferriere, 2001] *Ferriere K. M.* The interstellar environment of our galaxy // *Reviews Modern Physics*. 2001. V. 73. P. 1031–1066.
- [Frank et al., 2003] *Frank A., Poludnenko A., Gardiner T., Lebedev V., Drake R.* Stellar outflows with new tools: advanced simulations and laboratory experiments // *RevMexAA (Serie de Conferencias)*. 2003. V. 15. P. 85–91.
- [Gregori et al., 1999] *Gregori G., Miniati F., Ryu D., Jones T. W.* Enhanced cloud disruption by magnetic field interaction // *Astrophysical J.* 1999. V. 527. P. L113–L116.
- [Hartquist, Dyson, 1988] *Hartquist T. W., Dyson J. E.* Mass pick-up in astronomical flows // *Astrophysics and Space Science (Ap&SS)*. 1988. V. 144. P. 615.
- [Hartquist et al., 1997] *Hartquist T. W., Dyson J. E., Williams R. J. R.* Mass Injection Rates Due To Supernovae and Cloud Evaporation in Starburst Superwinds // *Astrophysical J. (ApJ)*. 1997. V. 482. P. 182.
- [Hunt et al., 1988] *Hunt J., Wray A., Moin P.* Eddies, streams, and convergence zones in turbulent flows // *Proc. Summer Program Center for Turbulence Research (NASA Ames/Stanford Univ.)*. 1988. P. 193–208.
- [Klein et al., 1994] *Klein R., McKee C., Colella P.* On the hydrodynamics interaction of shock waves with interstellar clouds, 1. Nonradiative shocks in small clouds // *ApJ*. 1994. V. 420. P. 213–236.
- [McKee, 1988] *McKee C. F.* Supernova remnant shocks in an inhomogeneous interstellar medium // *Supernova Remnants and the Interstellar Medium*. 1988. IAU Coll. 101. P. 205–222.
- [McKee, Cowie, 1975] *McKee C. F., Cowie L. L.* The Interaction between the Blast Wave of Supernova Remnant and Interstellar Clouds // *Astrophysical J.* 1975. V. 195. P. 715–725.
- [Melioli et al., 2005] *Melioli C., de Gouveia Dal Pino E., Raga A.* Multidimensional hydrodynamical simulations of radiative cooling SNRs–clouds interactions: an application to starburst environments // *Astronomy and Astrophysics*. 2005. V. 443. P. 495–508.
- [Picone, Boris, 1988] *Picone J. M., Boris J. P.* Vorticity generation by shock propagation through bubbles in a gas // *J. Fluid Mech.* 1988. V. 189. P. 23–51.
- [Rybakin, 2013] *Rybakin B.* Modeling of III-D Problems of Gas Dynamics on Multiprocessing Computers and GPU // *Computers and Fluids*. 2013. V. 80. P. 403–407.
- [Rybakin, Goryachev, 2014] *Rybakin B., Goryachev V.* The supersonic shock wave interaction with low-density gas bubble // *Acta Astronautica*. 2014. V. 94. Iss. 2. P. 749–753.
- [Rybakin, Goryachev, 2015] *Rybakin B., Goryachev V.* Treatment and visual analysis of numerical simulation of supersonic flows with extensive output after parallel calculation // *10th Pacific Symp. Flow Visualization and Image Processing*. Naples, Italy, 15–18 June, 2015. P. 1–8.
- [Rybakin et al., 2014] *Rybakin B. P., Stamov L. I., Egorova E. V.* Accelerated solution of problems of combustion gas dynamics on GPUs // *Computers and Fluids*. 2014. V. 90. P. 164–171.
- [Stone, Norman, 1992] *Stone J. M., Norman M. L.* The three-dimensional interaction of a supernova remnant with an interstellar cloud // *ApJL*. 1992. V. 390. P. 17–19.
- [Toro, 2009] *Toro E. F.* *Riemann Solvers and Numerical Methods for Fluid Dynamics*. Springer-Verlag, 2009. 724 p.

[Truelove et al., 1998] *Truelove J. K., Klein R. I., Mckee Ch. F., Holliman J. H., Howell L. H., Greenough J. A., Woods D. T.* Self-Gravitational Hydrodynamics with Three-Dimensional Adaptive Mesh Refinement: Methodology and Applications to Molecular Cloud Collapse And Fragmentation // *Astrophysical J.* 1998. V. 495. P. 821–852.

MATHEMATICAL MODELING OF THE PROCESSES OF FRAGMENTATION AND ABLATION OF THE MOLECULAR CLOUDS BY STRONG SHOCK WAVE

B. P. Rybakin

Scientific Research Institute of System Analysis, Russian Academy of Science (SRISA RAS)
Lomonosov MSU

The results of computer simulation of a shock wave interaction with a molecular cloud in three-dimensional statement are presented in the paper. Clouds compression caused by shock waves, jets, and other processes are common in galaxies. As the result of such interaction the Molecular Clouds (MC) are either destroyed or the processes of stars and planetary systems formation are accelerated in them. Shock waves, which propagate in the Interstellar Media (ISM), are formed as a result of various astrophysical processes such as jets, supernova explosions, collisions of MC, etc. For simulation the three-dimensional parallel program has been constructed with sufficient effective resolution for the cloud diameter. The detailed analysis of the processes of molecular cloud ablation, deformation and fragmentation has been carried out. Formation of passed and reflected waves system, material compression and ablation has been studied particularly. Post-processor treatment of calculations results made it possible to find features of emerging flow — the formation of vortexes, erosion and ablation of molecular cloud substance.

Keywords: supersonic flows, shock waves, parallel calculations.

Rybakin Boris Petrovich — leading scientist, doctor of physical and mathematical science, rybakin@vip.niisi.ru

ПРИМЕНЕНИЕ СОВРЕМЕННЫХ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ТЕХНОЛОГИЙ ДЛЯ МОДЕЛИРОВАНИЯ ЗАДАЧ СЕЙСМИКИ И ГЕОФИЗИКИ

Н. И. Хохлов, И. Б. Петров

Московский физико-технический институт (МФТИ)

В работе рассматривается программный комплекс, созданный для моделирования задач распространения динамических волновых возмущений в твёрдых телах. Комплекс использует двумерные и трёхмерные структурные блочные сетки с наличием неоднородностей. Для численного интегрирования применены сеточно-характеристические и конечно-объёмные методы 2–4 порядка точности. Алгоритм распараллелен по технологиям MPI. Реализована низкоуровневая оптимизация кода с использованием потоковых SIMD инструкций центральных процессоров SSE и AVX. В системах с общей памятью алгоритм распараллелен по технологии OpenMP, а код — по технологии CUDA. Приведены результаты тестирования ускорения и примеры расчётов.

Исследование выполнено при поддержке Российского фонда фундаментальных исследований (проект № 15-37-20673 мол_а_вед) и гранта Президента РФ МК-3383.2014.9.

Ключевые слова: MPI, CUDA, OpenMP, SIMS, SSE, AVX, сеточно-характеристический метод, системы гиперболических уравнений, механика деформируемого твёрдого тела.

ВВЕДЕНИЕ

Численное моделирование распространения динамических волновых возмущений в твёрдых телах используется при решении широкого круга задач, в том числе, сейсморазведки, сейсмоки и сейсмостойкости. Математическое моделирование проводится в различных геологических средах, в том числе, в слоистых и в средах с наличием различных неоднородностей (например, трещин или каверн). Задачи такого рода представляются очень ресурсоёмкими с точки зрения вычислительных ресурсов. Область вычисления, как правило, представляет собой сейсмический куб с длиной ребра от 1 до 10 км. В то же время, неоднородности могут быть размером в несколько метров. При моделировании задач сейсмостойкости также приходится сталкиваться с множеством неоднородностей. Параметры зданий, такие как толщина стен, размеры проёмов и пр. намного меньше размеров расчётных областей, включающих в себя порой массивы породы более 10 км вдоль одного направления. Таким образом, расчётная сетка должна быть достаточно подробной, чтобы иметь возможность правильно выделить все неоднородности. Для достаточной точности расчёта и учёта большого числа неоднородностей требуется использование больших вычислительных сеток размерами зачастую до нескольких десятков миллиардов узлов.

В данной работе рассматривается программный комплекс, созданный для моделирования задач распространения динамических волновых возмущений в твёрдых телах. Комплекс использует двумерные и трёхмерные структурные блочные сетки с наличием неоднородностей. Для численного интегрирования применяются сеточно-характеристические [Голубев и др., 2013; Иванов и др., 1990; Магомедов, Холодов, 1969, 1988; Петров, Тормасов, 1990; Петров, Холодов,

Хохлов Николай Игоревич — старший научный сотрудник, кандидат физико-математических наук, k_h@inbox.ru

Петров Игорь Борисович — главный научный сотрудник, член-корреспондент Российской академии наук, petrov@mipt.ru

1984а, б; Петров и др., 1989; Холодов, Холодов, 2006] и конечно-объёмные [Рое, 1986] методы 2–4 порядка точности. Алгоритм распараллелен по различным технологиям написания параллельных приложений. В настоящее время достигнута эффективность распараллеливания до 70 % по технологии MPI при масштабировании до 16 тысяч вычислительных ядер. В системах с общей памятью алгоритм распараллелен с использованием технологии OpenMP, технологии CUDA, что обеспечивает 50-кратное ускорение по сравнению с одним ядром CPU. Программа может использовать несколько карточек в рамках одного хоста. В данной работе рассмотрены результаты одного и того же алгоритма при использовании различных технологий. Приведены тесты распараллеливания до 16 тысяч ядер CPU и 6 устройств CUDA. В качестве примера работы программного комплекса приводится расчёт сейсмического отклика от системы трещин.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассмотрим основные уравнения линейной динамической теории упругости, которым подчиняется состояние бесконечно малого объёма линейно-упругой среды:

$$\begin{cases} \rho \dot{v}_i = \nabla_j \sigma_{ij}, \\ \dot{\sigma}_{ij} = q_{ijkl} \dot{\epsilon}_{kl}. \end{cases} \quad (1)$$

Здесь ρ — плотность среды; v_i — компоненты вектора скорости смещения; σ_{ij} и ϵ_{kl} — компоненты тензоров напряжений Коши и деформации; ∇_j — ковариантная производная по j -й координате. Вид компонент тензора 4-го порядка q_{ijkl} определяется реологией среды. Для линейно-упругого случая они имеют вид:

$$q_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}).$$

В этом соотношении, которое обобщает закон Гука, λ и μ — параметры Ламе; δ_{ij} — символ Кронекера.

Первая строка в системе (1) представляет три уравнения движения, вторая — шесть реологических соотношений. Вектор искомых функций, состоящий из 9 компонент, имеет вид:

$$\mathbf{u} = \{v_1, v_2, v_3, \sigma_{11}, \sigma_{12}, \sigma_{13}, \sigma_{22}, \sigma_{23}, \sigma_{33}\}^T.$$

Отметим, что система динамики деформируемого твёрдого тела (1) может быть записана в матричном виде:

$$\frac{\partial \mathbf{u}}{\partial t} = \sum_{j=1}^3 \mathbf{A}_j \frac{\partial \mathbf{u}}{\partial x_j}, \quad (2)$$

где \mathbf{A}_j — матрицы размером 9×9 , (x_1, x_2, x_3) — ортонормированная система координат.

Для её решения используется сеточно-характеристический метод на криволинейных гексаэдральных расчётных сетках [Петров, Холодов, 1984а]. Исходная система уравнений (2) преобразуется к виду:

$$\begin{cases} \frac{\partial \mathbf{u}}{\partial t} = \sum_{j=1}^3 \tilde{\mathbf{A}}_j \frac{\partial \mathbf{u}}{\partial \xi_j}, \\ \tilde{\mathbf{A}}_j = \sum_{i=1}^3 \frac{\partial \xi_j}{\partial x_i} \mathbf{A}_i, \end{cases} \quad (3)$$

где (ξ_1, ξ_2, ξ_3) — направления расщепления в преобразованном пространстве. Система (3) распадается на три одномерных системы уравнений:

$$\frac{\partial \mathbf{u}}{\partial t} = \tilde{\mathbf{A}}_j \frac{\partial \mathbf{u}}{\partial \xi_j}, j=1, 2, 3. \quad (4)$$

Каждая из этих систем является гиперболической и обладает полным набором собственных векторов с действительными собственными значениями. Каждую из систем можно переписать в виде:

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{\Omega}_j^{-1} \mathbf{\Lambda}_j \mathbf{\Omega}_j \frac{\partial \mathbf{u}}{\partial \xi_j},$$

где матрица $\mathbf{\Omega}_j$ — матрица, составленная из собственных векторов; $\mathbf{\Lambda}_j$ — диагональная матрица. Для каждого шага расщепления для направления ξ_j матрица $\mathbf{\Lambda}_j$ имеет вид:

$$\mathbf{\Lambda}_j = \text{diag}\{c_1 l_j, -c_1 l_j, c_2 l_j, -c_2 l_j, c_2 l_j, -c_2 l_j, 0, 0, 0\}.$$

Здесь $\tilde{n}_1 = \sqrt{\frac{\lambda + 2\mu}{\rho}}$; $\tilde{n}_2 = \sqrt{\frac{\mu}{\rho}}$; $l_j = |w^j| = \sqrt{(w_1^j)^2 + (w_2^j)^2 + (w_3^j)^2}$; $w^j = \nabla \cdot \xi_j$. Для определения коэффициентов w^j вычисляется матрица, обратная к матрице Якоби. Её необходимо вычислять в каждой ячейке. Коэффициенты могут быть вычислены аналитически, в случае, когда преобразование задано $\xi = \xi(\mathbf{x})$, либо численно с использованием формулы второго порядка точности $(w_i^j)_m = ((\xi_j)_{m+1} - (\xi_j)_{m-1}) / \Delta x_i$.

После замены переменных $\mathbf{v} = \mathbf{\Omega} \mathbf{u}$ каждая из систем (4) распадается на девять независимых скалярных уравнений переноса (индекс j далее опускается, где это возможно):

$$\frac{\partial v}{\partial t} + \mathbf{\Lambda} \frac{\partial v}{\partial x} = 0.$$

Одномерные уравнения переноса решаются с помощью метода характеристик [Холодов, Холодов, 2006]. После того, как все компоненты \mathbf{v} перенесены, восстанавливается само решение:

$$\mathbf{u}^{n+1} = \mathbf{\Omega}^{-1} \mathbf{v}^{n+1}.$$

В программе реализованы схемы 2–4 порядка точности. В работе расчёты проводились с использованием схемы 4-го порядка точности:

$$\begin{aligned} v_m^{n+1} &= v_m^n - \sigma \left(\Delta_1 - \sigma \left(\Delta_2 - \sigma \left(\Delta_3 - \sigma \Delta_4 \right) \right) \right), \\ \Delta_1 &= \frac{1}{24} \left(-2v_{m+2}^n + 16v_{m+1}^n - 16v_{m-1}^n + 2v_{m-2}^n \right), \\ \Delta_2 &= \frac{1}{24} \left(-v_{m+2}^n + 16v_{m+1}^n - 30v_m^n + 16v_{m-1}^n - v_{m-2}^n \right), \\ \Delta_3 &= \frac{1}{24} \left(2v_{m+2}^n - 4v_{m+1}^n + 4v_m^n - 2v_{m-2}^n \right), \\ \Delta_4 &= \frac{1}{24} \left(v_{m+2}^n - 4v_{m+1}^n + 6v_m^n - 4v_{m-1}^n + v_{m-2}^n \right). \end{aligned}$$

Кроме того, используется сеточно-характеристический критерий монотонности. В случае положительного значения λ критерий монотонности имеет следующий вид:

$$\min \{v_m^n, v_{m-1}^n\} \leq v_m^{n+1} \leq \max \{v_m^n, v_{m-1}^n\}.$$

Для отрицательных значений λ он будет симметричен. В простейшей реализации в случае невыполнения данного критерия монотонности применяется корректировка решения вида:

$$v_m^{n+1} = \begin{cases} \max \{v_m^n, v_{m-1}^n\}, & v_m^{n+1} > \max \{v_m^n, v_{m-1}^n\}, \\ \min \{v_m^n, v_{m-1}^n\}, & v_m^{n+1} < \min \{v_m^n, v_{m-1}^n\}, \\ v_m^{n+1}, \min \{v_m^n, v_{m-1}^n\} \leq v_m^{n+1} \leq \max \{v_m^n, v_{m-1}^n\}. \end{cases}$$

Данный ограничитель сохраняет 4-й порядок в областях, где решение ведёт себя достаточно гладко (выполняется характеристический критерий). В случае больших градиентов решения порядок схемы понижается до 3-го.

Приведённый выше алгоритм позволяет находить решение во внутренних точках области. Для этого случая все характеристики, выходящие из расчётного узла на следующем временном слое, остаются лежать в области интегрирования. В граничных узлах ситуация иная. Одна характеристика выходит из расчётной области [Петров, Холодов, 19846]. В таком случае необходимо использовать граничные условия. Учёт граничных условий осуществляется практически одинаково для всех границ. Рассмотрим переход от инварианта Римана к естественным переменным в граничном узле:

$$\mathbf{u}^{n+1} = \Omega^{-1} \mathbf{v}^{n+1} = \sum_i v_i^{n+1} \mathbf{r}_i, \tag{5}$$

где \mathbf{r}_i — векторы столбцы матрицы Ω^{-1} , v_i^{n+1} — соответствующие компоненты вектора \mathbf{v}^{n+1} . Разобьём сумму (5) на две части. Для первого слагаемого оставим те члены, для которых соответствующие характеристики попадают внутрь расчётной области, обозначим множество индексов i как λ^n . У второго оставим те характеристики, которые выйдут за границы области интегрирования — λ^{out} . Тогда равенство (5) можно переписать в виде:

$$\mathbf{u}^{n+1} = \sum_i v_i^{n+1} \mathbf{r}_i = \sum_{i \in \lambda^n} v_i^{n+1} \mathbf{r}_i + \sum_{i \in \lambda^{out}} v_i^{n+1} \mathbf{r}_i = \sum_{i \in \lambda^n} v_i^{n+1} \mathbf{r}_i + \Omega^{out} \mathbf{v}^{out} = \mathbf{u}^{in} + \Omega^{out} \mathbf{v}^{out}, \tag{6}$$

где Ω^{out} — прямоугольная матрица, составленная из правых собственных векторов, соответствующих характеристикам, которые выходят за область интегрирования; \mathbf{v}^{out} — вектор, составленный из соответствующих компонент инварианта Римана. Вектор \mathbf{u}^{in} может быть найден решением соответствующих уравнений переноса (4). Остаётся найти вектор \mathbf{v}^{out} .

Граничное условие в линейном случае может быть представлено в виде:

$$\mathbf{B} \mathbf{u}^{n+1} = \mathbf{b}, \tag{7}$$

где \mathbf{B} — прямоугольная матрица; \mathbf{b} — известный вектор. Подставив (7) в (6), получим запись граничных условий в общем виде:

$$\mathbf{B}(\mathbf{u}^{in} + \Omega^{out} \mathbf{v}^{out}) = \mathbf{b}.$$

Отсюда можно выразить вектор \mathbf{v}^{out} :

$$\mathbf{v}^{out} = -(\mathbf{B} \Omega^{out})^{-1} (\mathbf{B} \mathbf{u}^{in} - \mathbf{b}).$$

С учётом найденных выходных значений инвариантов Римана и выражения (6) получим:

$$\mathbf{u}^{n+1} = \mathbf{u}^{in} - \Omega^{out} (\mathbf{B} \Omega^{out})^{-1} (\mathbf{B} \mathbf{u}^{in} - \mathbf{b}).$$

Рассмотрим несколько вариантов граничных условий. Если на границе задана поверхностная плотность силы воздействия \mathbf{f} , то на границе выполнено условие $\mathbf{T}\mathbf{n} = \mathbf{f}$, где \mathbf{T} — тензор напряжений; \mathbf{n} — внешняя нормаль. В случае свободной границы внешняя сила равна нулю. Корректировка значений в граничных узлах имеет вид:

$$\left. \begin{aligned} \mathbf{v}^{n+1} &= \mathbf{v}^n - \frac{1}{\rho c_s} \mathbf{z} + \left[\frac{1}{c_s} - \frac{1}{c_p} \right] (\mathbf{z} \cdot \mathbf{n}) \mathbf{n}, \\ \mathbf{T}^{n+1} &= \mathbf{T}^n - [\mathbf{z} \otimes \mathbf{n} + \mathbf{n} \otimes \mathbf{z}] - \frac{\mathbf{z} \cdot \mathbf{n}}{\lambda + 2\mu} (\lambda \mathbf{I} - 2(\lambda + \mu)(\mathbf{n} \otimes \mathbf{n})), \end{aligned} \right\} \quad (8)$$

где вектор $\mathbf{z} = \mathbf{T}^n \mathbf{n} - \mathbf{f}$; \mathbf{I} — единичный тензор.

Если на границе задана внешняя скорость \mathbf{V} , то на границе выполнено условие $\mathbf{v} = \mathbf{V}$. В случае если скорость нулевая — граница жёстко закреплена. Корректировка значений в граничных узлах имеет вид:

$$\begin{aligned} \mathbf{v}^{n+1} &= \mathbf{V}, \\ \mathbf{T}^{n+1} &= \mathbf{T}^n - \rho \left\{ (\mathbf{n} \cdot \mathbf{z}) [(c_s - 2c_p - c_3)(\mathbf{n} \otimes \mathbf{n}) + c_3 \mathbf{I}] + c_s (\mathbf{n} \otimes \mathbf{z} + \mathbf{z} \otimes \mathbf{n}) \right\}, \\ \mathbf{z} &= \mathbf{v}^n - \mathbf{V}, \\ c_3 &= \sqrt{\frac{\lambda^2}{\rho(\lambda + 2\mu)}}. \end{aligned}$$

Дополнительно рассмотрим смешанные граничные условия. В таком случае задана нормальная составляющая скорости и тангенциальная составляющая силы. В случае отсутствия силы трения между телами, тангенциальная составляющая силы равна нулю. Если обозначить тангенциальную составляющую касательных сил \mathbf{f}_τ , а нормальную составляющую скорости V_n , то граничные условия можно записать в виде:

$$\mathbf{T}\mathbf{n} = \mathbf{f}, \quad \mathbf{v} \cdot \mathbf{n} = V_n, \quad \mathbf{f} = \mathbf{f}_\tau - (\mathbf{f}_\tau \cdot \mathbf{n}) \mathbf{n} + (\mathbf{f} \cdot \mathbf{n}) \mathbf{n}.$$

Корректировка будет осуществляться по формуле (8), в которой \mathbf{f} задаётся выражением:

$$\mathbf{f} = \mathbf{f}_\tau + \left[\rho c_p V_n + (\mathbf{T}^n \mathbf{n} - \rho c_p \mathbf{v}^n - \mathbf{f}_\tau) \cdot \mathbf{n} \right] \mathbf{n}. \quad (9)$$

В случае контакта двух тел производится корректировка соответствующих граничных узлов. В данной работе использовались условия свободного скольжения. При такой постановке у контактирующих тел совпадают нормальные компоненты скорости на границе, нормальные компоненты силы компенсируют друг друга, и тангенциальные компоненты силы равны нулю. Обозначим контактирующие тела индексами a и b , тогда условия контакта можно записать в виде:

$$\mathbf{v}_a \cdot \mathbf{n} = \mathbf{v}_b \cdot \mathbf{n} = V_n, \quad f_p^a = -f_p^b, \quad \mathbf{f}_\tau^a = \mathbf{f}_\tau^b = 0. \quad (10)$$

Тогда из выражения (9) с учётом соотношения (10) можно найти силу для корректировки узлов в каждом из тел:

$$\left. \begin{aligned} f_p^a &= \rho c_p V_p + (\mathbf{T}_a^m \mathbf{n} - \rho c_p \mathbf{v}_a^m) \cdot \mathbf{n}, \\ f_p^b &= \rho c_p V_p + (-\mathbf{T}_b^m \mathbf{n} - \rho c_p \mathbf{v}_b^m) \cdot \mathbf{n}, \\ V_p &= \frac{1}{2} \left[\mathbf{v}_a^m + \tilde{\mathbf{v}}_b^m - \frac{1}{\rho c_p} (\mathbf{T}_a^m - \mathbf{T}_b^m) \cdot \mathbf{n} \right] \cdot \mathbf{n}. \end{aligned} \right\} \quad (11)$$

ОПТИМИЗАЦИЯ РАБОТЫ ОДНОПОТОЧНОЙ ВЕРСИИ АЛГОРИТМА

Рассмотрим базовую организацию работы расчётного алгоритма. Используются явные сеточные методы, т. е. переход от одного шага по времени к следующему происходит путём обхода всей расчётной сетки и вычисления значения в новом узле, исходя из значения в исходном узле и некоторой его окрестности. Обход сетки организуется путём двух вложенных циклов. Исследование такой реализации показало, при достаточно большом размере расчётной сетки возникает определённое число промахов кэша процессора, что ведёт к увеличению времени работы программы. Путём изменения алгоритма обхода расчётной сетки, а также за счёт использования дополнительных массивов памяти и принудительного кеширования данных, удалось оптимизировать работу с памятью, что позволило в ряде случаев ускорить работы алгоритма на больших расчётных сетках. Величина ускорения зависит от модели центрального процессора.

На рис. 1 приведена зависимость числа тактов процессора, затраченных на перенос одного узла расчётной сетки в зависимости от размера сетки. Показаны результаты для базовой (Not optimized) и оптимизированной версии (Optimized). Размер сетки указан по одной из осей, тест проводился на двумерной задаче. В тесте участвовали следующие процессоры и компиляторы (указан размер кэша L2):

- Intel Xeon 5110, 4 МБ, icc-11.1;
- Intel Core2 Quad Q8200, 2 МБ, gcc-4.4.4.

При компиляции тестовых программ использовались следующие опции компиляторов: gcc-O3; icc-O3. Как видно из графиков, в зависимости от модели процессора ускорение может достигать 5 раз.

Данный алгоритм также был доработан для использования потокового SIMD-расширения центрального процессора (SSE). SSE включает в архитектуру несколько расширенных регистров (8 на современных процессорах) размером 128, либо 256 бит в случае технологии AVX. Преимущество в производительности достигается в том случае, если производится одна и та же последовательность действий над разными данными. В таком случае блоком SSE осуществляется распараллеливание вычислительного процесса между данными.

Для возможности оперирования данными в SSE регистрах потребовалось изменение формата хранения расчётной сетки. Чтобы произвести какие-либо операции, используя SSE, требуется поместить в регистр необходимые расчётные данные, в случае когда эта операция занимает много времени все преимущество потоковых инструкций теряется. В связи с этим возникает необходимость возможности быстрой загрузки и выгрузки данных в SSE регистры. Изначально данные хранились в виде массива структур, или AOS (Array of structures), при таком подходе необходимо по одному числу типа float или double копировать данные в SSE, затем производить вычисление и копировать данные обратно. При этом переменные, предназначенные для использования в одном SSE регистре, будут расположены в памяти не последовательно.

Для ускорения процесса необходимо расположить все компоненты структуры в памяти последовательно и затем подгружать их по необходимости в SSE регистры. Был произведён переход к хранению структуры массивов, или SOA (Structure of arrays), что позволило быстрее осуществлять копирование в SSE регистры. На рис. 2 приведено время работы различных версий алгоритма на одних и тех же данных.

Результаты приведены для базовой реализации, версии оптимизированной по памяти и использующей систему хранения SOA (Cache-friendly), а также двух

вариантов потоковых инструкций SSE и AVX. Размер расчётной сетки, использованной в тесте — 4096×4096 . Тестирование проводилось для различных вариантов процессоров и компиляторов, все процессоры поддерживали инструкции AVX:

- Intel Xeon E5-2690, gcc-4.4.6;
- Intel Xeon E5-2690, icc-12.1.0;
- AMD Opteron 6276, gcc-4.7.2.

Также менялся тип используемых данных: float и double.

По результатам тестирования видно ускорение до семи раз при использовании чисел одинарной точности и до двух раз при числах с двойной точностью. В зависимости от версии компилятора и процессора результат несколько отличается, в частности, в ряде случаев имеется автоматическая векторизация. Скорость работы алгоритма, использующего AVX инструкции, практически не отличается от скорости работы алгоритма, построенного на использовании инструкций SSE. Эта особенность связана с тем, что скорость работы памяти при загрузке данных в регистры AVX такая же, как и у SSE.

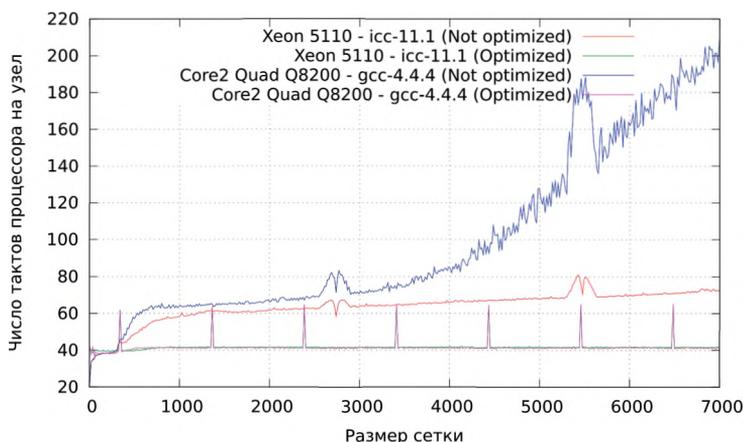


Рис. 1 Зависимость времени переноса одного узла разностной схемы от размера сетки

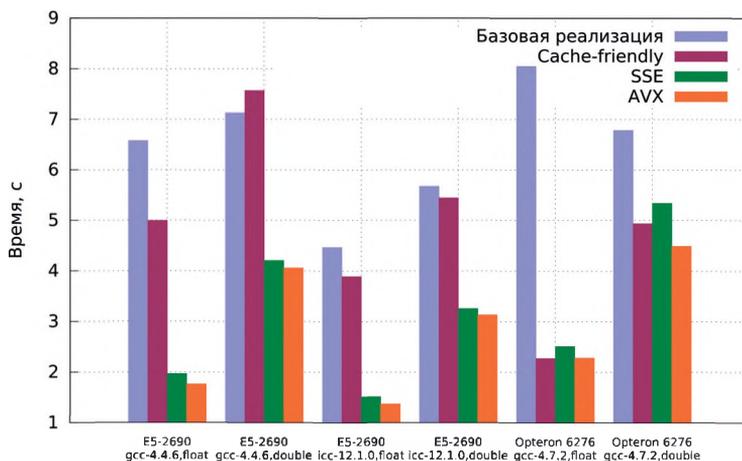


Рис. 2. Время работы различных вариантов оптимизации

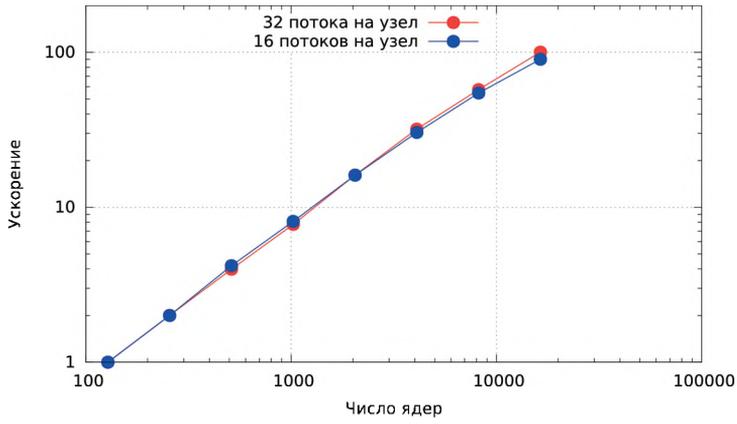


Рис. 3. Масштабируемость параллельного алгоритма

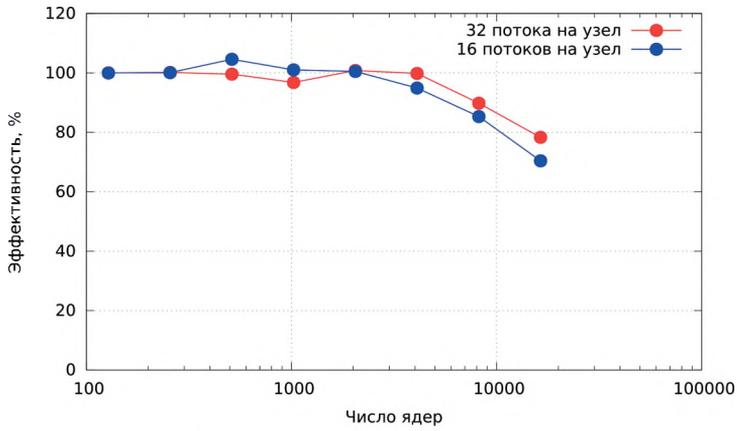


Рис. 4. Эффективность параллельного алгоритма

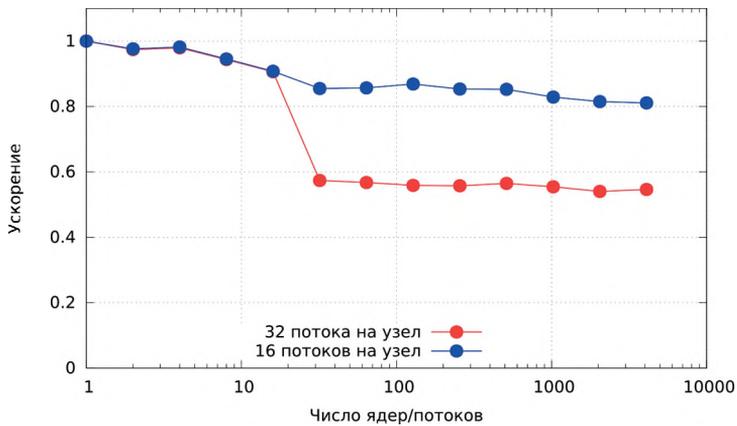


Рис. 5. Ускорение параллельного алгоритма при увеличении размера задачи

РАСПАРАЛЛЕЛИВАНИЕ MPI

Для работы в системах с распределённой памятью программный комплекс распараллелен по технологии MPI с использованием стандартных алгоритмов декомпозиции расчётной области и обмена приграничными ячейками. Одно из основных требований к алгоритму — работа на большом числе вычислительных ядер (тысячи) для обеспечения приемлемого времени расчёта. На рис. 3 приведены результаты тестирования масштабирования алгоритма при увеличении числа расчётных ядер от 128 до 16 384.

В тесте использовалась расчётная сетка размером $1000 \times 100 \times 01000$ узлов. Тестирование проводилось на кластере HECToR. Данный суперкомпьютер состоит из 2816 вычислительных узлов. Каждый из узлов оснащён двумя процессорами 16-core AMD Opteron 2.3GHz Interlagos. Оперативная память составляет 32 ГБ на узел. Ускорение составляет до 100 раз при увеличении числа вычислительных ядер в 128 раз. На рис. 4 приведены графики эффективности работы параллельного алгоритма.

Тестирование показало, что эффективность составляет около 70...80 %, что является хорошим результатом для такого количества вычислительных ядер. Результаты несколько отличаются при запуске 16 и 32 потоков на один узел суперкомпьютера. Это связано с особенностями работы вычислительных узлов. В случае, когда на каждом из ядер работает один процесс и интенсивно использует шину памяти, происходит провал в скорости её работы. В результате, если на одном узле запускать по 32 потока, происходит провал в скорости работы с памятью. Данная особенность относится только к текущему вычислительному комплексу и не является особенностью работы алгоритма.

Ещё один тест — масштабируемость размеров задачи при одновременном увеличении числа ядер. В данном тесте на каждое вычислительное ядро приходится одинаковое число узлов расчётной сетки. На рис. 5 приведены результаты такого тестирования.

Тестирование проводилось при изменении числа вычислительных ядер от 1 до 4096. На графике виден характерный провал при расчёте в 32 потока, связанный с особенностями работы вычислительного комплекса, описанными выше. Максимальный размер сетки в данном тесте — около 33 млрд узлов. Как можно видеть, алгоритм показывает хорошее ускорение и может вести расчёт на больших задачах.

РАСПАРАЛЛЕЛИВАНИЕ OPENMP

Алгоритм также распараллелен в системах с общей памятью при использовании технологии OpenMP. В данном направлении работа только начата и при увеличении числа ядер от 1 до 32 алгоритм показывает эффективность около 60 %. Столь низкие показатели связаны с недостаточно эффективной работой с памятью, вычислительным ядрам приходится много обращаться к областям памяти у соседних ядер, что вызывает замедление расчёта. Ведутся работы по повышению эффективности работы с технологией OpenMP.

РАСПАРАЛЛЕЛИВАНИЕ CUDA

Алгоритм также был распараллелен на графических GPGPU-процессорах NVidia при использовании технологии CUDA. Потребовалось полное переписывание части расчётного модуля под данную архитектуру.

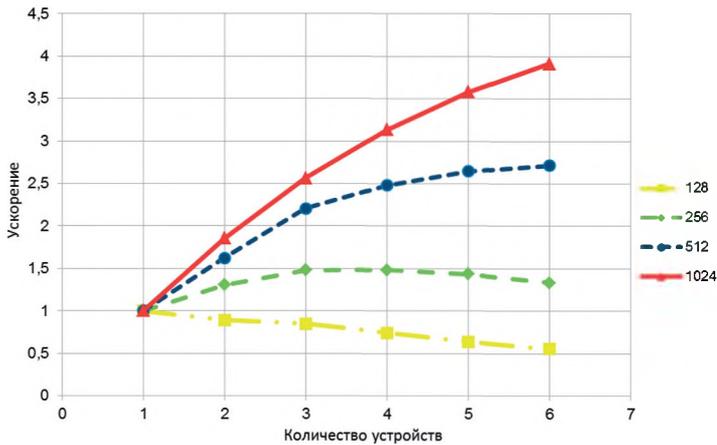


Рис. 6. Распараллеливание на нескольких карточках nVidia

На текущий момент достигнуто примерно 80-кратное ускорение по сравнению с одним ядром CPU при использовании числа с одинарной точностью, и 30-кратное — с двойной. В тестах использовались карточки nVidia Tesla s2050 и процессоры Intel Xeon X5670 2.7 ГГц. Текущая реализация алгоритма позволяет использовать несколько GPU карточек в рамках одного расчётного узла. Это сделано средствами технологии CUDA. На рис. 6 приведён тест ускорения при увеличении числа карточек от 1 до 6, в зависимости от размера расчётной сетки.

При небольших размерах сетки алгоритм не даёт прироста скорости работы при увеличении числа устройств. Это связано с большими накладными расходами на взаимодействие между устройствами. Однако на больших расчётных сетках реализация даёт ускорение до 4 раз при увеличении числа устройств в 6 раз. Ведутся работы по переносу алгоритма на гибридные архитектуры MPI+CUDA.

РАСЧЁТ ВОЛНОВОГО ОТКЛИКА ОТ ТРЕЩИНОВАТОГО КОЛЛЕКТОРА В ТРЁХМЕРНОЙ ПОСТАНОВКЕ

В качестве примера работы программного комплекса приведён расчёт сейсмического отклика от кластера пустых и флюидонасыщенных трещин.

Для описания вертикальных трещин применялась параллелепипедная расчётная сетка. Моделировался кластер из 31 параллельной трещины протяжённостью 3000 м и высотой 100 м, расположенных на расстоянии 100 м друг от друга. Он находился на глубине 2000 м от дневной поверхности. Вмещающий массив представлял собой параллелепипед ($10 \times 10 \times 3$ км) со следующими упругими характеристиками: скорость продольных волн — 4500 м/с, скорость поперечных волн — 2250 м/с, плотность среды — 2500 кг/м^3 (рис. 7). В качестве источника возмущения использовалась продольная упругая волна, распространяющаяся от дневной поверхности под углом 5° к вертикали в плоскости OXZ .

В ходе численного расчёта производится вычисление полного вектора скорости и тензора напряжения в последовательные моменты времени в каждом узле расчётной сетки, что позволяет визуализировать волновую картину вдоль произвольного сечения области (рис. 8). Необходимо отметить, однако, что при полевых измерениях имелась возможность регистрации лишь колебаний дневной поверхности.

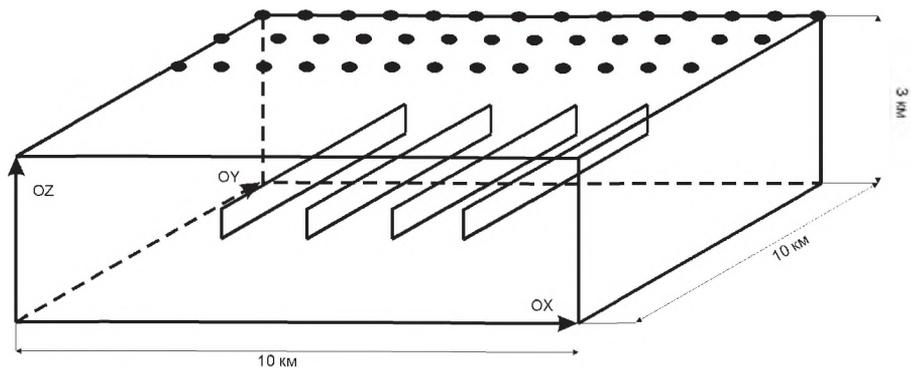
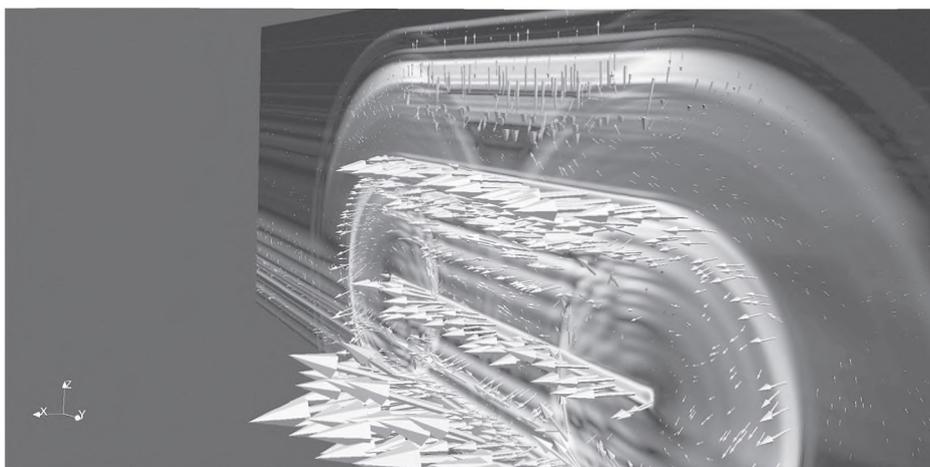
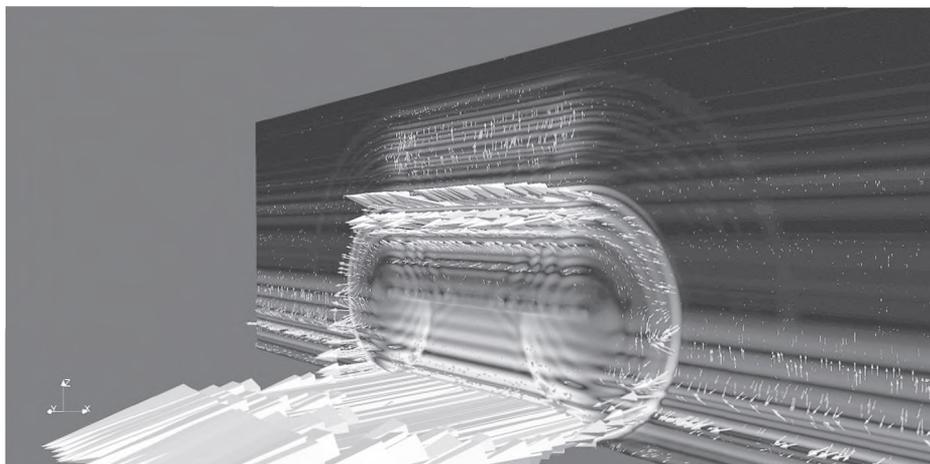


Рис. 7. Геометрия расчётной области в задаче о кластере вертикальных трещин



a



б

Рис. 8. Сечение расчётной области плоскостью OYZ . Насыщение вертикальных трещин: пустота (слева) и флюид (справа). Модуль скорости в градации серого цвета и вектор скорости

В ходе расчёта на верхней границе параллелепипеда выполнялась запись полного вектора скорости в местах расстановки сейсмоприёмников. Расстановка представляла собой площадное покрытие в виде квадратной сетки с шагом 50 м и осями, параллельными OX и OY .

Шаг сетки составлял 5 м вдоль оси OZ и 10 м вдоль осей OX и OY , и она содержала 600 млн узлов. Запуск программы производился на вычислительной системе с 30 расчётными ядрами. Поскольку при таком объёме вычислений, несмотря на параллельность алгоритма, одно выполнение программы занимало примерно сутки, было решено задавать заглубленное начальное возмущение на расстоянии 1700 м от дневной поверхности (над кластером).

Анализ результатов численных расчётов позволяет установить некоторые особенности динамических процессов, происходящих в трёхмерной трещиноватой среде. Вне зависимости от вида насыщения трещин происходит формирование как продольных сейсмических волн отклика, так и поперечных волн. Одной из интересных особенностей процесса является генерация периодического «шлейфа» упругих колебаний, сопровождающего основные фазы отклика и, по-видимому, вызванного межтрещинным взаимодействием. По всей ширине кластера инициируется плоский волновой фронт, а от его краёв начинают распространяться сферические волны. При этом при насыщении трещин флюидом интенсивность проходящего через трещиноватый кластер возмущения намного выше, чем при отсутствии заполнителя.

ЛИТЕРАТУРА

- [Голубев и др., 2013] *Голубев В. И., Петров И. Б., Хохлов Н. И.* Численное моделирование сейсмической активности сеточно-характеристическим методом // Журн. вычислительной математики и математической физики. 2013. Т. 53. № 10. С. 1709–1720.
- [Иванов и др., 1990] *Иванов В. Д., Петров И. Б., Суворова Ю. В.* Расчёт волновых процессов в наследственных вязкоупругих средах // Механика композитных материалов. 1990. № 3. С. 447–450.
- [Магомедов, Холодов, 1969] *Магомедов К. М., Холодов А. С.* О построении разностных схем для уравнений гиперболического типа на основе характеристических соотношений // Журн. вычислительной математики и математической физики. 1969. Т. 9. № 2. С. 383–396.
- [Магомедов, Холодов, 1988] *Магомедов К. М., Холодов А. С.* Сеточно-характеристические численные методы. М.: Наука, 1988. 288 с.
- [Петров, Тормасов, 1990] *Петров И. Б., Тормасов А. Г.* О численном исследовании трехмерных задач обтекания волнами сжатия препятствия или полости в упруго-пластическом полупространстве // Доклады АН СССР. 1990. Т. 314. № 4. С. 817–820.
- [Петров, Холодов, 1984а] *Петров И. Б., Холодов А. С.* О регуляризации разрывных численных решений уравнений гиперболического типа // Журн. вычислительной математики и математической физики, 1984. Т. 24. № 8. С. 1172–1188.
- [Петров, Холодов, 1984б] *Петров И. Б., Холодов А. С.* Численное исследование некоторых динамических задач механики деформируемого твёрдого тела сеточно-характеристическим методом // Журн. вычислительной математики и математической физики. 1984. Т. 24. Вып. 5. С. 722–739.
- [Петров и др., 1989] *Петров И. Б., Тормасов А. Г., Холодов А. С.* О численном изучении нестационарных процессов в деформируемых средах многослойной структуры // Известия АН СССР. Сер. МТТ. 1989. № 4. С. 89–95.

[Холодов, Холодов, 2006] *Холодов А. С., Холодов Я. А.* О критериях монотонности разностных схем для уравнений гиперболического типа // Журн. вычислительной математики и математической физики. 2006. Т. 46. № 9. С. 1638–1667.

[Roe, 1986] *Roe P. L.* Characteristic-Based Schemes for the Euler Equations // Annual Review of Fluid Mechanics. 1986. No. 18. P. 337–365.

APPLICATION OF MODERN HIGH PERFORMANCE TECHNOLOGIES FOR SEISMIC AND GEOPHYSICS MODELING

N. I. Khokhlov, I. B. Petrov

Moscow Institute of Physics and Technology (MIPT)

The numerical simulation of seismic wave propagation presents an essential part of the exploration work the oil industry has to address on a day-to-day basis. The application presented here allows the simulation of the propagation of dynamic wave disturbances in various geological mediums, including layered mediums and mediums with different inhomogeneities (e.g. cracks or cavities). The type of numerical problem solved in these simulations is extremely compute-intensive and mandates large-scale HPC systems. The area of the ground medium simulated in these computations is typically a seismic cube with an edge length in the order of 1km to 10km. At the same time, inhomogeneities can be as small as a few meters in dimension. Therefore, the computational grid needs to be adequately small to be able to properly capture all inhomogeneities. Obtaining sufficient accuracy and specifying a large number of irregularities requires the use of very large computational grids; real-world computations use grids with several billion nodes.

This paper describes SEISMO, a new numerical seismology code that models the propagation of seismic waves in ground media with inhomogeneities on 2D and 3D structured and unstructured grids, and outlines its performance on a range of current HPC platforms. The mathematical model used in SEISMO is based on the theory of linear elasticity. The finite-difference and finite-volume methods are then used to solve the resulting hyperbolic partial differential equations (PDEs). SEISMO implements the Total Variation Diminishing (TVD) discretisation scheme. The code is parallelised using MPI, with low-level optimisations of the algorithm on SIMD CPUs using SSE and AVX instructions; it currently runs at 70% parallel efficiency when scaled up to 16k cores. The aim is to improve on this by exploring the benefits of using hybrid and accelerator programming approaches. The core algorithm of the application was ported to GPU accelerators using NVidia CUDA.

Keywords: MPI, CUDA, OpenMP, SIMS, SSE, AVX, grid-characteristic method, hyperbolic systems, elastic equation.

Khokhlov Nikolay Igorevich — senior scientist, PhD, k_h@inbox.ru

Petrov Igor Borisovich — project chief designer, corresponding member of the Russian Academy of Sciences, petrov@mipt.ru

К ВОПРОСУ ОБ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ ИНФРАСТРУКТУРЕ, А ТАКЖЕ О ВОЗМОЖНОСТИ ПРОВЕДЕНИЯ УЧРЕЖДЕНИЯМИ ФАНО ИССЛЕДОВАНИЙ В ОБЛАСТИ ИНТЕНСИВНОЙ ОБРАБОТКИ ДАННЫХ С ЦЕЛЬЮ ИЗВЛЕЧЕНИЯ НОВОГО ЗНАНИЯ

Л. Н. Щур

Научный центр Российской академии наук в Черноголовке (НЦЧ РАН)

Приводится анализ возможной роли информационно-телекоммуникационной (ИКТ) инфраструктуры Федерального агентства научных организаций (ФАНО) в развитии фундаментальных исследований. Особый акцент сделан на роль ИКТ в области интенсивной обработки Больших данных (Data-Intensive Scientific Discovery). Причём, как с целью извлечения нового знания, так и для разработки методов и механизмов осуществления такой деятельности научными коллективами. Мы предлагаем подход к выработке предложений по развитию ИКТ инфраструктуры и разработке программно-аппаратных решений для предоставления возможности конечному пользователю работы с Большими данными (BD) непосредственно с его рабочего места.

Ключевые слова: информационно-телекоммуникационная инфраструктура, хранение и передача данных большого объёма, интенсивная обработка больших данных, распределённый программно-аппаратный комплекс, интерфейс пользователя.

ВВЕДЕНИЕ

Начало 1990-х гг. в мире отмечено активным развитием компьютерных сетей. Наиболее активную роль при этом играли учреждения науки и образования. Инициатива ученых привела к разработке администрацией США документа, который предопределил политику и технологии создания современных компьютерных сетей, а также их насыщения программными продуктами и информационными технологиями (подробнее см. статью [Щур, 2012]). В нашей стране создание телекоммуникационной инфраструктуры было итогом активности сотрудников учреждений Российской академии наук, Курчатовского института и университетов, среди которых следует особо отметить Московский государственный университет имени М. В. Ломоносова (МГУ), Ярославский государственный университет им. П. Г. Демидова (ЯрГУ) и Новосибирский национальный исследовательский государственный Университет (НГУ). Положительную роль в процессе создания сети для науки и образования сыграл в 1990-е гг. фонд Сороса (Институт «Открытое общество»). Создание и развитие в России компьютерных сетей для науки и образования происходило на средства, выделяемые Российским фондом фундаментальных исследований (РФФИ) и Министерством науки РФ. С 1997 по 2013 г. развитие компьютерных сетей РАН проводилось в рамках программы целевых расходов Президиума РАН «Создание средств телекоммуникаций и их обслуживание», позднее преобразованной в программу РАН «Высокопроизводительные вычислительные системы и телекоммуникации».

Щур Лев Николаевич — заместитель председателя НЦЧ РАН, заведующий отделом прикладных сетевых исследований НЦЧ РАН, заведующий базовой кафедрой Научно-исследовательского университета Высшая школа экономики «Прикладные информационно-коммуникационные средства и системы», ведущий научный сотрудник Института теоретической физики им. Л. Д. Ландау РАН, профессор, доктор физико-математических наук, e-mail: Shchur@chg.ru

К настоящему времени на балансе научных учреждений, подведомственных ФАНО, имеется серьёзная информационно-телекоммуникационная (ИКТ) инфраструктура. Ниже мы приводим её основные характеристики.

1. ИКТ ИНФРАСТРУКТУРА УЧРЕЖДЕНИЙ ФАНО — ТЕХНИЧЕСКИЕ АСПЕКТЫ

ИКТ инфраструктура учреждений ФАНО с технической точки зрения состоит:

- из магистральных каналов связи;
- коммуникационного оборудования, обеспечивающего связность сетей учреждений ФАНО с сетями других ведомств науки и образования РФ;
- коммуникационного оборудования, обеспечивающего связность с научными сетями зарубежных стран;
- систем хранения данных;
- систем обработки данных.

2. ФУНКЦИИ ИКТ ИНФРАСТРУКТУРЫ УЧРЕЖДЕНИЙ ФАНО

С функциональной точки зрения ИКТ инфраструктура учреждений ФАНО обеспечивает, например:

- обмен данными большого объёма между учреждениями ФАНО;
- распределённые научные семинары и конференции учреждений ФАНО;
- фундаментальные научные исследования с использованием ИКТ инфраструктуры учреждений ФАНО.

3. ОСОБЕННОСТЬ ИКТ ИНФРАСТРУКТУРЫ УЧРЕЖДЕНИЙ ФАНО

Функциональность ИКТ инфраструктуры учреждений ФАНО обеспечивается за счёт наличия третьей составляющей ИКТ инфраструктуры, которая включает:

- проведение особой сетевой политики учреждениями ФАНО;
- обеспечение особой сетевой политики взаимодействия с сетями науки и образования;
- проведение фундаментальных научных исследований с использованием ИКТ инфраструктуры учреждений ФАНО.

4. ИКТ ИНФРАСТРУКТУРА УЧРЕЖДЕНИЙ ФАНО И ОБРАБОТКА БОЛЬШИХ ДАННЫХ

Ожидаемый прорыв в развитии фундаментальных исследований связывается с реализацией методологии «Большие данные» (Big Data, BD) с использованием подхода «Наука с интенсивным использованием данных» (Data Intensive Scientific Discovery, DISD). Всё более признанным становится мнение, что мы находимся на пороге возникновения новой парадигмы научных исследований в дополнении к уже устоявшимся подходам получения нового знания путём эксперимента, теории и компьютерного моделирования [Gray, 2009]. Был проведён анализ информационных ресурсов, которые потенциально могут быть использованы для выполнения DISD. Для этого были проанализированы ресурсы 29 университетов РФ, получивших статус НИУ, учреждений Отделения физических наук (ОФН) РАН, нескольких ведущих университетов и государственных научных лабораторий

США; рассмотрены работы по модернизации инфраструктуры университетов США и Европы с целью предоставления возможности работы с ВД, как сервиса для научных сотрудников и аспирантов.

Проведённые предварительные исследования показали, что интуитивное и распространённое мнение о том, что «...в Интернете много доступных научных данных», оказалось сильно преувеличенным.

На сегодняшний день формальное определение Больших данных основано на четырёх V, т. е. (Very) Big Data — это следующий набор признаков V⁴:

- *Volume* — большой объём данных;
- *Velocity* — большая скорость обработки потока данных;
- *Variety* — большое разнообразие данных;
- *Veracity* — большая достоверность данных.

Именно на этом наборе данных и можно ожидать, что исследования с интенсивной обработкой таких данных могут привести к получению принципиально нового знания.

4.1. Большие данные научных и образовательных учреждений РФ

Информационные ресурсы учреждений Отделения физических наук (ОФН) РАН состоят в основном, из электронных библиотек, а также ограниченного набора данных по исследованиям в области астрофизики, исследований космоса и метеорологии. Среди Национальных исследовательских университетов России (НИУ РФ) лишь половина имеет некоторые информационные научные ресурсы (МГУ, Высшая школа экономики (ВШЭ), Нижегородский государственный университет имени Н. И. Лобачевского (ННГУ), НГУ, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (ИТМО)). В основном это данные в области социологии.

Более благоприятная картина в области полных текстов научных статей. Двадцать два российских научных журнала имеют полные тексты статей — это «Письма в ЖЭТФ», «Успехи физических наук», «Труды МИАН», «Прикладная механика и теоретическая физика», «Проблемы передачи информации», «Математические заметки», «Журнал экспериментальной и теоретической физики (ЖЭТФ)». Причём два первых журнала предоставляют в свободном доступе полные тексты статей всех томов с начала основания. Такая комбинация полноты и открытости является уникальной, и выделяет эти журналы в общемировом ИКТ пространстве.

Заметим, что имеется большое количество научных баз данных и электронных хранилищ научных публикаций, но доступ к ним ограничен специальными соглашениями финансового и организационного порядка.

4.2. Большие данные научных и образовательных учреждений зарубежья

Зарубежные научные учреждения содержат в основном локальную, с закрытым доступом, информацию для сотрудников и студентов. Открытая часть информации содержит лишь демонстрационный набор данных, в основном, с целью популяризации научных исследований и создания престижа учреждения. Исключение составляют только открытые данные университета Калтех (Калифорния, США), который предоставляет открытый доступ к научным данным в области генома,

исследования мозга, геодинамики и исследования космоса. Эти данные, однако, не удовлетворяют требованиям по объёму, предъявляемым к Большим данным.

Среди зарубежных научных журналов пятнадцать имеют электронные версии всех статей за всю историю журналов, например, *Philosophical Transactions*, *Nature*, *Physical Review*, *Chemical Reviews*, *Science*. Эти тексты находятся в закрытом доступе и для ознакомления с ними требуется специальное соглашение.

4.3. Тенденции в области научных исследований с интенсивной обработкой данных

Что может дать доступ к Большим данным? Что можно из них извлечь?

Наиболее яркий и общеизвестный пример — это обработка данных с Большого адронного коллайдера (БАК), которая привела к экспериментальному обнаружению бозона Хиггса, элементарной частицы с энергией 125 ГэВ. Это один из самых крупных научных проектов. Для обработки экспериментальных данных была внедрена технология грид, которая объединила 140 вычислительных центров в 35 странах мира. Затраты на создание инфраструктуры по передаче, хранению и обработке данных составили более половины затрат на поиски самого бозона Хиггса. Технология Грид развёрнута поверх научных компьютерных сетей стран-участниц и использует специальные каналы связи между странами. Это первый пример из разряда Big Data, он удовлетворяет 1-му, 2-му и 4-му признакам Больших данных, приведённых выше. Кроме того, это пример построения мега-установки и создания специализированной ИКТ инфраструктуры для фундаментальной науки.

Второй пример, на этот раз в полной мере удовлетворяющий всем четырём признакам Больших данных, реализован фирмой ИБМ. В научно-исследовательской лаборатории Accelerated Discovery Lab (ADLab) фирмы в Алмадене (Калифорния, США) развернут программно-аппаратный комплекс для хранения и обработки Больших данных, а также для проведения фундаментальных исследований в области интенсивной обработки данных [Pease, 2015]. Возможности программно-аппаратного комплекса «ИБМ Ватсон» были продемонстрированы широкой общественности в ходе телевикторины, которую «ИБМ Ватсон» выиграл у многократных её чемпионов. Кроме того, «ИБМ Ватсон» принадлежит серьёзное достижение в области поиска веществ, влияющих на рост злокачественных клеток, — за полгода исследований с использованием Больших данных, хранящихся в центре Алмадена, были обнаружены два фермента, влияющих на протеин p53, ответственный за процесс активации роста злокачественных клеток. Обнаружение новых ферментов с такими свойствами делается всеми лабораториями мира с частотой один фермент в год. Заметим, что и в этом случае был создан специализированный программно-аппаратный комплекс для решения задачи. Создание комплекса такого размера требует значительных средств и высококвалифицированного персонала.

4.4. Опыт отдела прикладных сетевых исследований НЦЧ РАН

Работы по созданию и развитию современной информационной, телекоммуникационной и вычислительной инфраструктуры в Научном центре РАН в Черноголовке начались в конце 1991 г. по инициативе группы научных сотрудников. В 1995 г. эта работа была формализована созданием лаборатории компьютерного обеспечения научных исследований, преобразованной в 2001 г. в отдел

прикладных сетевых исследований (ОПСИ). ОПСИ также ведёт фундаментальные научные исследования в области информатики, прикладной математики и вычислительной физики, как в рамках государственного задания, так и по грантам РФФИ, Российского научного фонда (РНФ), РАН и Минобрнауки [Шикота, 2013].

В настоящее время ОПСИ имеет следующие элементы ИКТ инфраструктуры:

- сеть ChANT — опорная оптоволоконная сеть, объединяющая учреждения ФАНО в Черноголовке на скорости 1-10 гигабит в секунду;
- телекоммуникационное оборудование;
- библиотечный сервер — доступ к каталогам БЕН РАН, БНЦ РАН, библиотечным сервисам [Шикота и др., 2011];
- сервер свободно распространяемого программного обеспечения — архив программного обеспечения, используемого учреждениями ФАНО [Щур и др., 2010];
- вычислительный кластер WALL — кластер из 12 узлов;
- гибридный вычислительный кластер Manticore — кластер из 4 узлов (8 процессоров, 80 физических ядер) с восемью графическими ускорителями;
- облачная инфраструктура «Тучка» — пробная система из трёх серверов;
- серверы хранения данных — около 0,5 Пбайт дискового пространства;
- система проведения распределённых видеоконференций ВидеоГрид — масштабируемый программно-аппаратный комплекс [Щур и др., 2012];
- распределённая информационная система совместной работы научных и образовательных коллективов Компфиз [Шикота и др., 2012].

Эта инфраструктура может быть использована для организации исследований в области интенсивной обработки больших данных.

5. ПРЕДЛОЖЕНИЯ ПО РАЗВИТИЮ ИКТ ИНФРАСТРУКТУРЫ УЧРЕЖДЕНИЙ ФАНО И ПО ПОДГОТОВКЕ К ВНЕДРЕНИЮ ПРОГРАММНО-АППАРАТНЫХ РЕСУРСОВ С ЦЕЛЬЮ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Мы делаем предложения по инициированию комплекса работ, цель которых состоит:

- в использовании имеющейся ИКТ инфраструктуры учреждений ФАНО для проведения фундаментальных исследований в областях с Data-Intensive Scientific Discovery (DISD);
- разработке программы научных исследований в области DISD для её выполнения силами научных коллективов учреждений ФАНО;
- разработке программы мероприятий по внедрению сервисов работы с Большими данными.

Неотложного решения требуют проблемы бюджетного финансирования поддержки телекоммуникационной инфраструктуры, и особо — магистральных каналов.

6. МОДЕРНИЗАЦИЯ ИКТ ИНФРАСТРУКТУРЫ С ЦЕЛЬЮ ВНЕДРЕНИЯ СИСТЕМ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В ПРОЦЕСС ПРОВЕДЕНИЯ ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Эта задача может быть решена в рамках государственного задания учреждениями ФАНО, имеющими на своём балансе соответствующие элементы ИКТ инфраструктуры. Как показывают успешные примеры технологий грид (LHC-Grid) и IBM Watson, задача создания программно-аппаратных комплексов для научных исследований в области работы с Большими данными требует серьёзных финансовых вложений и привлечения высококвалифицированных специалистов. В этих примерах такое стало возможно, в первом случае, за счёт консолидации финансовых средств десятков государств и объединения в рамках проектов и рабочих групп около 5000 исследователей, а во втором — благодаря организации специальной лаборатории ADLab и привлечения IT Department IBM для создания и поддержки программно-аппаратного комплекса. Эти примеры показывают, что учреждениям ФАНО под силу создать распределённый программно-аппаратный комплекс, который предоставит всем научным сотрудникам и аспирантам возможность работы с Большими данными со своего рабочего места.

6.1. О программе научных исследований в области DIS

Необходимо создать программу по поддержке научных исследований в области проведения фундаментальных исследований с интенсивной обработкой данных. Такая программа должна учитывать приоритетные направления фундаментальных исследований, проводимых учреждениями ФАНО.

6.2. О программе мероприятий по разработке методов и систем обработки Больших данных

Необходимо создать программу мероприятий по разработке методов, систем, алгоритмов и программного обеспечения для работы с Большими данными. Вышеприведённые примеры демонстрируют важность совместной работы специалистов по информатике, прикладной математике и предметных областей для успешного выполнения такой задачи.

В подготовке записки приняли участие сотрудники отдела прикладных сетевых исследований НЦЧ РАН, которым автор выражает глубокую благодарность. С их помощью проведено сканирование общедоступных ресурсов ведущих научных учреждений и Высших учебных институтов (ВУЗов) РФ, научных программ и проектов в области обработки больших данных.

Работа выполнена в рамках государственного задания НЦЧ РАН.

ЛИТЕРАТУРА

- [Шикота, 2013] Шикота С. К. К задаче развития инфокоммуникационной инфраструктуры научного центра // Информац. об-во. 2013. № 1-2. С. 102–111.
- [Шикота и др., 2011] Шикота С. К., Крашаков С. А., Шур Л. Н. Единая система библиотечных сервисов научного центра // Высокие технологии, фундаментальные исследования, экономика. 2011. № 2. С. 127–129.

- [Шикота и др., 2012] *Шикота С. К., Меньшутин А. Ю., Щур Л. Н.* Программно-аппаратная платформа для научных исследований в области вычислительной физики // Программная инженерия. 2012. № 9. С. 16–22.
- [Щур, 2012] *Щур Л. Н.* Роль инфокоммуникационных технологий в развитии процесса глобализации научных исследований // Информац. об-во. 2012. № 5. С. 16–24.
- [Щур и др., 2010] *Щур Л. Н., Крашаков С. А., Меньшутин А. Ю.* 20-летний опыт использования СПО в Научном центре в Черноголовке для организации научной работы // Тез. докл. Всерос. конф. «Свободное программное обеспечение-2010». СПб., 2010. С. 19.
- [Щур и др., 2011] *Щур Л. Н., Меньшутин А. Ю., Шикота С. К.* Инфокоммуникационное обеспечение исследовательского центра: задачи и инфраструктура // Информац. об-во. 2011. № 6. С. 58–65.
- [Hey et al., 2009] *Hey T., Tansley S., Tolle K.* The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, Washington: Microsoft research, 2009. 287 p.
- [Pease, 2015] *Pease D. A.* The IBM Research Accelerated Discovery Lab: Objectives and Experience // Proc. 17th Intern. Conf. “Data Analytics and Management in Data Intensive Domains” (DAMDID/RCDL’2015). Obninsk, Russia, October 13–16, 2015. P. 201–202.

ON THE INFO-COMMUNICATION INFRASTRUCTURE OF THE INSTITUTIONS OF RUSSIAN ACADEMY OF SCIENCES, AND ON THE INITIATION OF RESEARCH IN THE DATA-INTENSIVE SCIENTIFIC DISCOVERY DOMAIN

L. N. Shchur

Science Center in Chernogolovka Russian Academy of Sciences (SCCh RAS)

We present analysis of the possible role of info-communication (IT) infrastructure of the institution of Russian Academy of Sciences in the fundamental research. We emphasise how IT infrastructure can be transformed to met requirements of Data-Intensive Scientific Discovery (DISD). It is two fold goal: first, it is new knowledge extraction, and, second, it is development of new methods and tools to give scientists possibility to work in DISD environment.

Keywords: info-communication infrastructure, storage and transmission of big data, data-intensive scientific discovery, distributed information systems, user interface

Shchur Lev Nikolaevich — deputy chairman of SCCh RAS, head of department SCCh RAS, chair of department “Applied Info-communication tools and systems”, leading scientist Landau Institute for Theoretical Physics, professor, doctor of physical and mathematical sciences, e-mail: Shchur@chg.ru

ВЫЧИСЛИТЕЛЬНЫЕ
ТЕХНОЛОГИИ
В ЕСТЕСТВЕННЫХ
НАУКАХ

**МЕТОДЫ
СУПЕРКОМПЬЮТЕРНОГО
МОДЕЛИРОВАНИЯ**

3

