



ВЕСТНИК

ЮЖНО-УРАЛЬСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА

2017
Т. 6, № 1

ISSN 2305-9052

СЕРИЯ

«ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И ИНФОРМАТИКА»

Решением ВАК включен в Перечень научных изданий,
в которых должны быть опубликованы результаты диссертаций
на соискание ученых степеней кандидата и доктора наук

Учредитель — Федеральное государственное автономное образовательное учреждение
высшего образования «Южно-Уральский государственный университет»
(национальный исследовательский университет)

Тематика журнала:

- Вычислительная математика и численные методы
- Математическое программирование
- Распознавание образов
- Вычислительные методы линейной алгебры
- Решение обратных и некорректно поставленных задач
- Доказательные вычисления
- Численное решение дифференциальных и интегральных уравнений
- Исследование операций
- Теория игр
- Теория аппроксимации
- Информатика
- Математическое и программное обеспечение высокопроизводительных вычислительных систем
- Системное программирование
- Перспективные многопроцессорные архитектуры
- Облачные вычисления
- Технология программирования
- Машинная графика
- Интернет-технологии
- Системы электронного обучения
- Технологии обработки баз данных и знаний
- Интеллектуальный анализ данных

Редакционная коллегия

С.М. Абдуллаев, д.г.н., проф.
А.В. Панюков, д.ф.-м.н., проф.
А.С. Порозов, *техн. секретарь*
Л.Б. Соколинский, д.ф.-м.н., проф., *отв. редактор*
В.П. Танана, д.ф.-м.н., проф., *зам. отв. редактора*
М.Л. Цымблер, к.ф.-м.н., доц., *отв. секретарь*

Редакционный совет

А. Андреяк, PhD, профессор (Германия)
В.И. Бердышев, д.ф.-м.н., акад. РАН, *председатель*

В.В. Воеводин, д.ф.-м.н., чл.-кор. РАН
Дж. Донгарра, PhD, профессор (США)
С.В. Зыкин, д.т.н., профессор
Д. Маллманн, PhD, профессор (Германия)
А.Н. Томилин, д.ф.-м.н., профессор
В.Е. Третьяков, д.ф.-м.н., чл.-кор. РАН
В.И. Ухоботов, д.ф.-м.н., профессор
В.Н. Ушаков, д.ф.-м.н., чл.-кор. РАН
М.Ю. Хачай, д.ф.-м.н., профессор
П. Шумяцки, PhD, профессор (Бразилия)
Е. Ямазаки, PhD, профессор (Бразилия)



BULLETIN

OF THE SOUTH URAL STATE UNIVERSITY **2017**
vol. 6, no. 1

SERIES

**“COMPUTATIONAL
MATHEMATICS AND SOFTWARE
ENGINEERING”**

ISSN 2305-9052

**Vestnik Yuzhno-Ural'skogo Gosudarstvennogo Universiteta.
Seriya “Vychislitel'naya Matematika i Informatika”**

South Ural State University

The scope of the journal:

- Numerical analysis and methods
- Mathematical optimization
- Pattern recognition
- Numerical methods of linear algebra
- Reverse and ill-posed problems solution
- Computer-assisted proofs
- Numerical solutions of differential and integral equations
- Operations research
- Game theory
- Approximation theory
- Computer science
- High performance computing
- System software
- Advanced multiprocessor architectures
- Cloud computing
- Software engineering
- Computer graphics
- Internet technologies
- E-learning
- Database processing
- Data mining

Editorial Board

S.M. Abdullaev, South Ural State University (Chelyabinsk, Russia)
A.V. Panyukov, South Ural State University (Chelyabinsk, Russia)
A.S. Porozov, South Ural State University (Chelyabinsk, Russia)
L.B. Sokolinsky, South Ural State University (Chelyabinsk, Russia)
V.P. Tanana, South Ural State University (Chelyabinsk, Russia)
M.L. Zymbler, South Ural State University (Chelyabinsk, Russia)

Editorial Council

A. Andrzejak, Heidelberg University (Germany)
V.I. Berdyshev, Institute of Mathematics and Mechanics, Ural Branch of the RAS (Yekaterinburg, Russia)
J. Dongarra, University of Tennessee (USA)
M.Yu. Khachay, Institute of Mathematics and Mechanics, Ural Branch of the RAS (Yekaterinburg, Russia)
D. Mallmann, Julich Supercomputing Centre (Germany)
P. Shumyatsky, University of Brasilia (Brazil)
A.N. Tomilin, Institute for System Programming of the RAS (Moscow, Russia)
V.E. Tretyakov, Ural Federal University (Yekaterinburg, Russia)
V.I. Ukhobotov, Chelyabinsk State University (Chelyabinsk, Russia)
V.N. Ushakov, Institute of Mathematics and Mechanics, Ural Branch of the RAS (Yekaterinburg, Russia)
V.V. Voevodin, Lomonosov Moscow State University (Moscow, Russia)
Y. Yamazaki, Federal University of Pelotas (Brazil)
S.V. Zykin, Sobolev Institute of Mathematics, Siberian Branch of the RAS (Omsk, Russia)

Содержание

Вычислительная математика

АЛГОРИТМ ПОСТРОЕНИЯ ИНТЕГРАЛЬНОГО ИНДИКАТОРА КАЧЕСТВА СЛОЖНОЙ СИСТЕМЫ ДЛЯ РЯДА ПОСЛЕДОВАТЕЛЬНЫХ НАБЛЮДЕНИЙ Т.В. Жгун	5
МОДИФИКАЦИЯ МЕТОДА МАТРИЧНЫХ ПУЧКОВ, ИСПОЛЬЗУЮЩАЯ СОВМЕСТНОЕ ОЦЕНИВАНИЕ ПОЛЮСОВ СИГНАЛА И ОБРАТНЫХ К НИМ О.Л. Ибряева, Д.Д. Салов	26
ЭКСПЕРИМЕНТАЛЬНОЕ СРАВНЕНИЕ АЛГОРИТМОВ В ПАРАЛЛЕЛЬНОМ МЕТОДЕ ВЛОЖЕННЫХ СЕЧЕНИЙ А.Ю. Пирова, Н.Ю. Кудрявцев, И.Б. Мееров	38

Геоинформатика

ИССЛЕДОВАНИЕ ВЛИЯНИЯ СТАЦИОНАРНЫХ ТЕЧЕНИЙ НА ДИНАМИЧЕСКИЕ ПРОЦЕССЫ И ЭВОЛЮЦИЮ ЗАГРЯЗНЕНИЙ В АЗОВСКОМ МОРЕ Л.В. Черкесов, Т.Я. Шульга	56
--	----

Информатика, вычислительная техника и управление

HYBRID COMPUTER SYSTEM PROGRAMMING TECHNOLOGY WITH ADAPTATION AND SCALING OF CALCULATIONS А.А. Gulenok, А.І. Dordopulo, І.І. Levin, V.A. Gudkov	73
---	----

Суперкомпьютерное моделирование

ПАРАЛЛЕЛЬНАЯ РЕАЛИЗАЦИЯ СТОХАСТИЧЕСКОЙ КЛЕТОЧНО-АВТОМАТНОЙ МОДЕЛИ РЕКОМБИНАЦИИ ЭЛЕКТРОНОВ И ДЫРОК В 2D и 3D НЕОДНОРОДНЫХ ПОЛУПРОВОДНИКАХ К.К. Сабельфельд, А.Е. Киреева	87
--	----

Contents

Computational Mathematics

- AN ALGORITHM FOR CONSTRUCTING INTEGRAL QUALITY INDICATOR OF
COMPLEX SYSTEMS FOR A SEQUENCE OF OBSERVATIONS
T.V. Zhgun 5
- MODIFICATION OF THE MATRIX PENCIL METHOD USING A COMBINED
EVALUATION OF SIGNAL POLES AND THEIR INVERSES
O.L. Ibryaeva, D.D. Salov 26
- EXPERIMENTAL EVALUATION OF ALGORITHMS IN THE PARALLEL MULTILEVEL
NESTED DISSECTION METHOD
A.Yu. Pirova, N.Yu. Kudriavtsev, I.B. Meyerov 38

Geoinformatics

- STUDY OF STATIONARY CURRENTS FOR DYNAMIC PROCESSES AND ADMIXTURES
OF POLLUTION IN THE SEA OF AZOV
L.V. Cherkesov, T.Ya. Shul'ga 56

Computer Science, Engineering and Control

- HYBRID COMPUTER SYSTEM PROGRAMMING TECHNOLOGY WITH ADAPTATION
AND SCALING OF CALCULATIONS
A.A. Gulenok, A.I. Dordopulo, I.I. Levin, V.A. Gudkov 73

Supercomputer Simulation

- PARALLEL IMPLEMENTATION OF STOCHASTIC CELLULAR AUTOMATA MODEL OF
ELECTRON-HOLE RECOMBINATION IN 2D AND 3D HETEROGENEOUS
SEMICONDUCTORS
K.K. Sabelfeld, A.E. Kireeva 87

АЛГОРИТМ ПОСТРОЕНИЯ ИНТЕГРАЛЬНОГО ИНДИКАТОРА КАЧЕСТВА СЛОЖНОЙ СИСТЕМЫ ДЛЯ РЯДА ПОСЛЕДОВАТЕЛЬНЫХ НАБЛЮЖДЕНИЙ

© 2017 г. Т.В. Жгун

Новгородский государственный университет имени Ярослава Мудрого
(173003 Великий Новгород, ул. Большая Санкт-Петербургская, д. 41),

E-mail: Tatyana.Zhgyn@novsu.ru

Поступила в редакцию: 26.10.2016

В статье предлагается алгоритм построения интегральной характеристики изменения качества системы на основании регистрируемых измерений, который обеспечивает решение задачи выделения сигнала в многомерном массиве данных в условиях априорной неопределенности о свойствах сигнала на основании задаваемого отношения сигнал/шум. Построение латентной интегральной характеристики изменения качества системы на основе статистических показателей для ряда последовательных наблюдений производится на основе метода главных компонент с учетом наличия шума в измеряемых данных (ОСП-алгоритм). В отличие от классического метода главных компонент, где информативность вычисленной интегральной характеристики задается априорно и обеспечивается выбором числа главных компонент, в предлагаемом алгоритме информативность решения оценивается апостериорно на основании дисперсионного критерия и выбранного параметра отношения сигнал/шум. С помощью предложенного алгоритма построены интегральные индикаторы качества жизни субъектов Российской Федерации за 2007–2014 годы.

Ключевые слова: качество системы управления, интегральная характеристика качества, изменение характеристики качества, интегральные индикаторы качества жизни, шум измеряемых данных, отношение сигнал/шум, метод главных компонент, информативность метода главных компонент.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Жгун Т.В. Алгоритм построения интегрального индикатора качества сложной системы для ряда последовательных наблюдений // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. Т. 6, № 1. С. 5–25. DOI: 10.14529/cmse170101.

Введение

Интегральные характеристики качества системы используются при решении задач управления для совершенствования механизмов управления на основании сопоставления интегральных индикаторов рассматриваемых объектов и рейтингов объектов, определяемых их интегральными характеристиками, и являются показателем степени достижения цели управления. Согласно данным ООН, к 2011 году существовало 290 индексов, разработанных для ранжирования или комплексной оценки стран в соответствии с некими экономическими, политическими, социальными или экологическими критериями. Обзор феноменального роста количества применяемых для комплексной оценки стран сводных индексов, проведенного Бандура [13–15], показал, что только 26 из них (9%) было сформировано до 1991 года, 132 (46%) — с 1991 года по 2006 год, а в период 2007–2011 годы было сформировано 130 сводных индексов.

Стремительное увеличение числа композитных индексов является явным признаком их важности в разработке государственной политики и оперативной значимости в макроэкономике в целом. Все основные международные организации, такие как Организация экономического сотрудничества и развития (OECD), Европейский Союз, Всемирный экономический форум или Международный Валютный Фонд конструируют составные

показатели в самых различных областях [15, 23]. Общей целью большинства из этих показателей является ранжирование стран и их сравнительный анализ по некоторой агрегированной мере [15, 16, 26, 27]. Возможность использовать в политических целях единый показатель, обобщающий сложные, плохо формализуемые процессы (качество жизни, финансовую устойчивость, рыночную политику и т. д.) является сильным искушением для заинтересованных сторон. Следовательно, улучшение методик построения таких показателей является очень важным вопросом исследования и с теоретической, и с практической точек зрения.

Создатели композитных индексов считают, что сводная статистика действительно адекватно описывает реальность, и такое выделение итога является чрезвычайно полезным в обеспечении интереса средств массовой информации и внимания политиков. С точки зрения противников составных характеристик, достаточно сформировать соответствующий набор показателей, и не имеет смысла идти далее, к получению интегральной характеристики. Их основными возражениями является произвольный характер определения весов, с помощью которых переменные объединяются в единую характеристику. Обсуждение плюсов и минусов составных показателей приводится в работах [16, 23].

Организация экономического сотрудничества и развития (*OECD*) и развития ведет непрерывную работу по совершенствованию методик построения композитных индексов [23–26]. В 2008 году *OECD* совместно с Объединенным исследовательским центром (*Joint Research Centre European Commission*) с подготовлен справочник [17], явившийся итогом многолетних исследований в этой области [23–26, 29], где для широкого круга заинтересованных лиц изложен набор технических принципов, которые должны помочь исследователям в формировании составных показателей. Основным методом агрегации данных авторы выбирают линейную свертку показателей, а основным инструментом построения сводных показателей — факторный анализ.

Для построения качественного интегрального показателя необходима, во-первых, тщательная теоретическая проработка теоретических аспектов измеряемого явления, ибо «то, что плохо определено, вероятно, будет плохо измерено» [23], во-вторых, качественный выбор данных, так как качество составных индикаторов в значительной степени зависит от качества базовых показателей, и в-третьих, адекватный инструмент для работы с многомерными данными.

Факторный анализ впервые был использован для объединения показателей в единый индекс при разработке индекса здоровья в работе Хайтауэра в 1978 году [19]. При вычислении Показателей социально-экономического статуса (*Socio-Economic Status Indices, SES*) метод главных компонент был принят в качестве стандарта метода построения, где вычисляемый индекс определяла проекция на первую главную компоненту [22, 30]. Эту же методику использовали Линдман и Селин при создании Индекса экологической устойчивости [21], Сомарриба и Пена при измерении качества жизни в Европе [28]. Среди отечественных исследований следует отметить работы С.А. Айвазяна [1, 2] по определению индекса качества жизни населения.

Однако первая главная компонента хорошо приближает моделируемую ситуацию, если максимальное собственное число ковариационной матрицы дает вклад не менее 70% в сумму всех собственных чисел. Такое соотношение выполняется, если рассматривается небольшое количество признаков (не более пяти), и одно из свойств системы явно доминирует над остальными. При описании социально-экономических систем число пе-

ременных значительно более пяти, а структура системы не допускает простого приближения. В качестве выхода из этой ситуации рассматривается согласно [2] понижение порога информативности до 55%, и деление исходной системы на подсистемы, описываемые меньшим количеством переменных. В рассмотренных выше исследованиях [19–22, 28, 30] вклад наибольшего собственного числа колебался от 13% до 38%, за исключением работы [28], рассматривавшей модельный пример, где эта цифра составила 56%. Авторы следовали рекомендациям [28], утверждавшим, что первая главная компонента дает удовлетворительные весовые показатели даже в тех случаях, когда наибольшее собственное число вносит небольшой вклад в сумму всех собственных чисел.

Исследователи Организации экономического сотрудничества и развития придерживаются иной точки зрения. Для формирования составного индикатора используют факторный анализ, где метод главных компонент применяется исключительно для извлечения факторов с тем, чтобы количество извлеченных факторов объясняло бы более 50% общей дисперсии. Значение композитного индекса в этом случае определяют только значимые нагрузки выбранных главных факторов после вращения. Де-факто именно эта методика становится стандартом при вычислении композитных индексов [20, 23, 24, 29]. Хотя авторы [23] обращают внимание, что различные методы извлечения основных компонент и различные способы вращения подразумевают различные значимые переменные, а значит, разные веса переменных при вычислении композита, следовательно, и различные значения вычисляемого интегрального индикатора. Кроме того, факторный анализ предполагает наличие достаточной корреляции между исходными переменными, что в некотором смысле противоречит идее полного описания исследуемого явления совокупностью независимых величин.

Следует также отметить еще одно обстоятельство. Методика определения весовых коэффициентов с помощью факторного анализа не может применяться для сравнения характеристик описываемых объектов в динамике, так как даже при фиксированных методах извлечения факторов и способа вращения, факторный анализ для разных наблюдений системы структуру главных факторов (значимые нагрузки) определяет по-разному, что делает бессмысленным межвременные сравнения [10].

Недостаточное удовлетворение исследователей качеством построенных композитных показателей основывается не только на методических проблемах построения, но и на недостаточном качестве используемых данных, так как фиксируемые статистическими организациями данные неизбежно содержат погрешности измерения. Тем не менее, именно статистические данные, содержащие неустраняемые погрешности, в настоящее время представляют собой наилучшие оценки имеющихся реальных величин в социальных системах [18, 23].

Получение точных характеристик объекта на основании однократного измерения, неизбежно содержащего неизвестную погрешность, не представляется возможным. Однако по серии таких измерений вычисление неизвестной характеристики вполне вероятно. Такую задачу, в частности, с успехом решает астрофотометрия, определяющая основные числовые параметры астрономических объектов не по однократному наблюдению (изображению), а по серии зашумленных изображений. Используя основные идеи, лежащие в основе астрофотометрии, будем рассматривать построение интегральной характеристики изменения качества сложной системы как решение задачи выделения полезного сигнала по серии наблюдений, содержащей описание неизвестного параметра (в

многомерном массиве зашумленных данных) в условиях априорной неопределенности о свойствах полезного сигнала на основании задаваемого отношения сигнал/шум.

Целью статьи является разработка методики построения интегральной характеристики изменения качества сложной системы на основании серии регистрируемых измерений с помощью метода главных компонент. В разделе 1 представлена постановка задачи вычисления интегральной характеристики изменения качества системы. В разделе 2 обсуждается вопрос определения информативности решения поставленной задачи при наличии шума в измеряемом сигнале. Раздел 3 посвящен алгоритму построения интегральной характеристики изменения качества системы на основании задаваемого отношения сигнал/шум. В последнем разделе с помощью предложенного алгоритма построены интегральные индикаторы изменения качества жизни субъектов Российской Федерации за 2007–2014 годы. В заключении делается вывод, что предлагаемая методика может быть использована для вычисления интегральных оценок изменения качества плохо формализуемых систем.

1. Постановка задачи вычисления интегральной характеристики изменения качества системы

Рассмотрим построение интегральной оценки системы из m объектов, для которой известны таблицы из n описаний объектов для ряда наблюдений $t = 1, 2, \dots$ — матрицы $A^t = \{a_{ij}\}_{i,j=1}^{n,m}$. Для каждого момента t вектор интегральных показателей имеет вид

$$q^t = A^t \cdot w^t, \quad (1)$$

где $q^t = \langle q^t_1, q^t_2, \dots, q^t_m \rangle^T$ — вектор интегральных индикаторов момента t ,

$w^t = \langle w^t_1, w^t_2, \dots, w^t_m \rangle^T$ — вектор весов показателей для момента t , A^t — матрица

предобработанных данных для момента t . Для построения интегрального индикатора качества системы требуется найти веса показателей w^t для каждого момента времени.

Будем рассматривать оцениваемый объект как сложную (не поддающуюся удовлетворительной формализации), большую (число состояний выше современных вычислительных возможностей) систему. Такими системами являются как каждый биологический объект, так и любая социальная система. Система доступна для наблюдения, и известно конечное, достаточно большое число регистрируемых с некоторой точностью численных характеристик этой системы в разные моменты времени. Значимость регистрируемых показателей для функционирования системы в общем случае неизвестна. Для решения задачи управления требуется дать мотивированную оценку каждого наблюдаемого объекта на всем промежутке наблюдений, т.е. вычислить в динамике интегральную характеристику качества системы. Численные характеристики системы предварительно подвергнуты унификации — приведению значений переменных на отрезок $[0, 1]$ по принципу: «чем больше, тем лучше».

Любое измерение, в том числе и статистическое, неизбежно связано с точностью измерительного прибора, поэтому результат измерения неизбежно содержит неустранимую ошибку. Построение интегральной характеристики системы можно рассматривать как задачу выделения полезного сигнала на фоне шумов в условиях отсутствия априор-

ной информации о статистических характеристиках аддитивного шума и полезного сигнала при наличии единственной реализации измеряемого процесса. Эта задача аналогична задаче восстановления цифровых изображений, искаженных белым гауссовским шумом. Метод главных компонент (МГК) позволяет выделить структуру в многомерном массиве данных и с успехом применяется для распознавания изображений и для шумоподавления.

Количественные характеристики конкретной системы, функционально связанные с ее структурными особенностями и условиями ее работы, зависят от отношения сигнала к шуму. Это отношение часто используется для количественной оценки эффективности различения сигналов в опико-электронных, телевизионных автоматических системах, в средствах контроля и диагностики.

ОСШ — отношение сигнал/шум (англ. *signal-to-noise ratio*, сокр. *SNR*) представляет собой отношение сигнала (а если точнее, суммы сигнала и шума) к шуму. Величина может быть вычислена либо как безразмерное отношение амплитуды сигнала к амплитуде шума $SNR = A_s / A_n$, либо в децибелах $SNR (dB) = 20 \cdot \log_{10} (A_s / A_n)$. Эта величина наиболее полно описывает качество воспроизведения сигнала в телевизионных системах, в системах мобильной связи, в астрофотометрии.

Выбор порогового значения отношения полезного сигнала к шуму фона, позволяющего различить сигнал на фоне шума, обоснован в [8]. Современные технические системы (и человеческий глаз) уверенно выделяют сигнал из шума, если уровень ОСШ системе составляет около 7 дБ (в безразмерных единицах — 2,2). Такое пороговое значение используется в фотометрии слабых объектов: при регистрации сигнала от тусклых звезд необходимо, чтобы отношение сигнал/шум превышало 2,2. Телевизионные системы могут восстановить из шума неповрежденную картинку, если отношение уровня сигнала к уровню шумов превосходит пороговое значение (SNR_{thr}) около 7дБ. Именно это значение $SNR_{thr} = 2,2$ и будет использоваться далее.

Статистические данные неизбежно содержат погрешность измерения. Любой результат, полученный на основании этих данных, будет содержать неустранимую ошибку. Переход к другому моменту времени означает изменение данных, которое при неизменной структуре системы вызвано как изменением ситуации, так и случайными ошибками. Метод главных компонент на основании различных для разных моментов значений собственных векторов и собственных значений описывает неизменную структуру системы. Следовательно, именно значения собственных чисел и собственных векторов будут тем неизменяющимся сигналом, который нужно распознать, т.е. по имеющимся реализациям выделить сигнал из зашумленных данных. Предположение о том, что при вариации входных данных у собственных чисел имеется общая тенденция, иллюстрирует рис.1, где представлены значения упорядоченных по убыванию собственных чисел для разных наблюдений. В среднем значении рассматриваемых величин хорошо просматривается тенденция (сигнал) и случайное отклонение от нее. Именно усреднение значений используется в астрофотографии для подавления шумов. Усреднение работает на основе предположения об абсолютно случайной природе шума. Соответственно, случайные отклонения от истинных данных будут последовательно снижаться по мере усреднения возрастающего числа наблюдений.

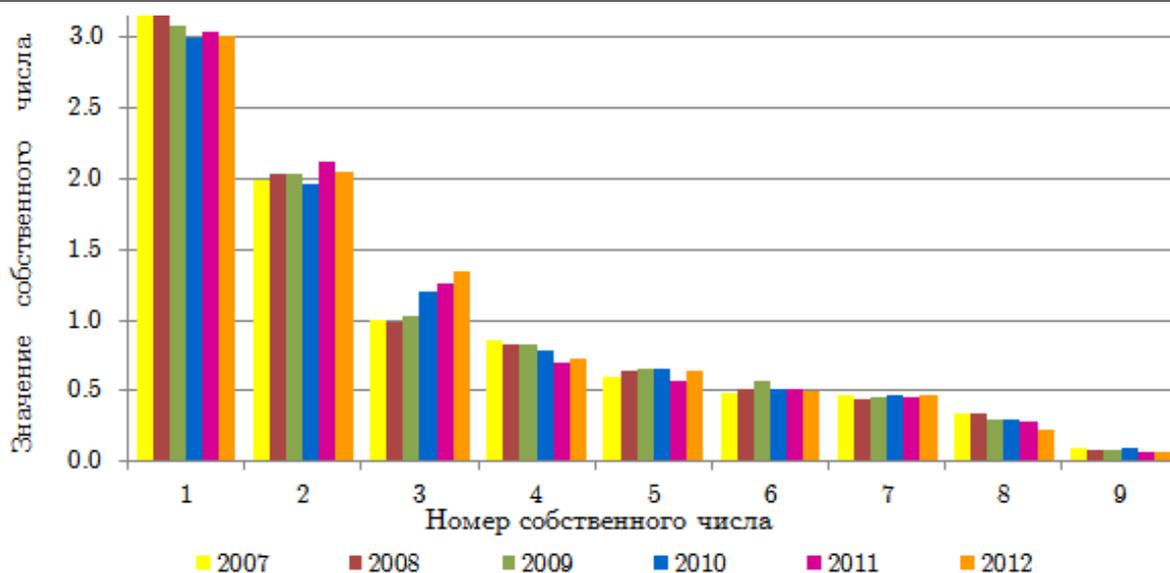


Рис. 1. Собственные значения ковариационной матрицы переменных для различных моментов наблюдения

Собственные векторы в методе главных компонент определяются с точностью до направления, в отличие от собственных чисел, определяемых однозначно. Среднее значение факторных нагрузок переменных зависит от выбранного направления и не может однозначно характеризовать сигнал. Следовательно, на основе вычисленных для различных наблюдений собственных векторов (упорядоченных в порядке убывания собственных чисел) нужно распознать случайные и неслучайные компоненты этих векторов и определить значения координат неслучайных переменных этих векторов.

Наличием неслучайного (т.е. значимого) вклада переменной в структуру главных компонент будем считать не большую величину факторной нагрузки после вращения, а инвариантность факторной нагрузки при возмущениях, признаком которой будет величина отношения сигнал/шум, определяемая средними значениями переменной (сигнал) и среднеквадратичным отклонением (шум). Если это отношение выше порогового значения, такую переменную считаем неслучайной — действующей. При невыполнении такого соотношения переменная характеризует шумовую составляющую сигнала и не участвует в дальнейшем рассмотрении. Для проверки этого условия нужно согласовать направления собственных векторов на основании некоторого критерия. После согласования направлений главных компонент и определения значимых переменных они, как и в факторном анализе, будут участвовать в дальнейшем рассмотрении, а незначимые переменные игнорируются (обнуляются). Модифицированные по такому принципу главные компоненты будем далее именовать эмпирическими главными компонентами (ЭГК).

2. Информативность интегральной характеристики изменения качества системы

В алгоритмах вычисления интегральной характеристики по ОСШ-методу [8–11] используется традиционное для МГК понятие информативности, которое определяет число главных компонент l , используемых для вычисления интегральной характеристики.

$$\gamma_{\sigma} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_l}{\lambda_1 + \lambda_2 + \dots + \lambda_n} \geq \theta. \quad (2)$$

Однако размерность пространства признаков в задачах вычисления интегральной характеристики качества сложной системы не является чересчур большой, и вычислительных проблем при определении собственных чисел и векторов не возникает. А качественное описание структуры системы требует либо всех главных компонент, либо их достаточно большого количества. Может оказаться, что ценная для конкретной задачи информация содержится как раз в направлениях с меньшей дисперсией. Например, при создании цифровой модели рельефа, которая строится по оцифрованным снимкам, искомым рельеф дают восьмая и девятая главные компоненты, а главные компоненты 12 и 13 в методе «Гусеница» свидетельствует о наличии в анализируемых данных периодичности с дробным периодом [5, 6].

Подходы к оценке числа главных компонент по необходимой доле объясненной дисперсии формально применимы всегда, однако неявно они предполагают, что нет разделения на «сигнал» и «шум», и любая заранее заданная точность имеет смысл. При разделии данных на полезный сигнал и шум задаваемая точность теряет смысл и требуется переопределить понятие информативности. Аналогично дисперсионной информативности согласно (2) можно определить SNR -информативность для выбранного числа эмпирических главных компонент N :

$$\gamma_{SNR} = \frac{S_{11} + S_{12} + \dots + S_{1N}}{S_{21} + S_{22} + \dots + S_{2N}}, \quad (3)$$

где S_{1k} — сумма величин ОСШ у действующих переменных k -ой ЭГК, S_{2k} — сумма ОСШ всех переменных k -ой ЭГК. Эта величина будет апостериорной оценкой (сверху) SNR -информативности. В отличие от дисперсионной информативности, SNR -информативность не может достигать 100% по логике построения. Информативность выбранной системы признаков определяется дисперсионной и SNR -информативностью:

$$\gamma = \gamma_{\sigma} \cdot \gamma_{SNR}. \quad (4)$$

3. Алгоритм построения интегральной характеристики изменения качества системы как решение задачи выделения полезного сигнала на фоне шумов

Приведем алгоритм построения интегральной характеристики изменения качества системы для ряда наблюдений на основании регистрируемых измерений, который обеспечивает решение задачи выделения сигнала в многомерном массиве данных в условиях априорной неопределенности о свойствах полезного сигнала на основании задаваемого отношения сигнал/шум. Основным параметром алгоритма является пороговое значение отношения полезного сигнала к шуму фона. По умолчанию используется пороговое значение $SNR_{thr} = 2,2$, применяемое в современных технических системах и позволяющее различить слабый сигнал на фоне шума. Возможно использование несколько более высокого порогового значения ОСШ, характеризующего именно слабый сигнал — до трех единиц. Если структура системы позволяет из содержательных соображений выделить подсистемы рассматриваемой системы, то переменные, описывающие систему, могут быть объединены в блоки и интегральная характеристика всей системы вычисляется через характеристики подсистем. Входными данными алгоритма являются:

- таблицы из n описаний m объектов для ряда наблюдений $t = 1, \dots, T$ — матрицы $A^t = \{a_{ij}^t\}_{i,j=1}^{n,m}$. Элемент матрицы a_{ij}^t — значение j -го показателя i -го объекта, вектор $a_i^t = (a_{i1}^t, \dots, a_{im}^t)$ — описание i -го объекта в момент t ;
- пороговое значение отношения сигнал/шум SNR_{thr} .

Алгоритм включает следующие шаги:

1 шаг. Подготовка данных

- обеспечение однородности выборки и унификация показателей.

Отсутствующие значения для отдельных наблюдений восстанавливаются интерполяцией по имеющимся данным. Устраняются те переменные, которые связаны друг с другом известной функциональной зависимостью (например, из двух переменных — количество мужчин и женщин в процентах — остается одна). Для прочих экспериментальных данных строгая коллинеарность практически не наблюдается, и переменные могут участвовать в решении задачи даже при высоких значениях парной корреляции.

Если исходный показатель связан с анализируемым интегральным свойством качества монотонной зависимостью, то при унифицировании исходных показателей (приведении их на отрезок $[0, 1]$ по принципу «чем больше, тем лучше») переменные x_{ij} для каждого момента наблюдения преобразовываются по правилу:

$$a_{ij} = s_j + (-1)^{s_j} \cdot \frac{x_{ij} - m_j}{M_j - m_j},$$

где $s_j = 0$, если оптимальное значение j -го показателя максимально и $s_j = 1$, если оптимальное значение j -го показателя минимально, m_j — наименьшее значение j -го показателя по всей выборке (глобальный минимум), M_j — наибольшее значение j -го показателя по всей выборке (глобальный максимум).

Если исходный показатель связан с анализируемым интегральным свойством качества жизни немонотонной зависимостью (т.е. внутри диапазона изменения данного показателя существует значение x_j^{opt} , при котором достигается наивысшее качество), то значение соответствующего унифицированного показателя рассчитывается по формуле:

$$a_{ij} = \left(1 - \frac{|x_{ij} - x_j^{opt}|}{\max((M_j - x_j^{opt}), (x_j^{opt} - m_j))} \right).$$

2 шаг. Определение априорных собственных векторов и собственных чисел для каждого наблюдения t

- центрирование (стандартизирование) унифицированных переменных: каждый элемент центрированной матрицы данных вычисляется по формуле:

$$z_{ij} = \frac{a_{ij} - \overline{a_j}}{s_j},$$

где $\overline{a_j} = \frac{1}{m} \sum_{i=1}^m a_{ij}$ — выборочная средняя, $s_j^2 = \frac{1}{m} \sum_{i=1}^m (x_{ij} - x_j)^2$ — выбо-

рочная дисперсия, т.е. s_j — выборочное среднеквадратичное отклонение;

- вычисление по центрированным значениям унифицированных показателей ковариационной матрицы, совпадающей в этом случае с матрицей корреляций;

– нахождение для каждого t собственных значений ковариационной (корреляционной) матрицы λ_i^t , $i = 1, \dots, n$, ранжирование их по убыванию;
 нахождение собственных векторов, соответствующих собственным числам λ_i^t для каждого наблюдения: $w^t_i = \langle w^t_{i1}, w^t_{i2}, \dots, w^t_{in} \rangle$, $t = 1, \dots, T$.

3 шаг. Определение эмпирических собственных чисел. Согласование направления собственных векторов и определение эмпирических главных компонент

- вычисление эмпирических значений собственных чисел как среднего значения собственных чисел для различных выборок $\lambda_i = \frac{1}{T} \sum_{t=1}^T \lambda_i^t$;
- согласование направлений собственных векторов (главных компонент) для разных наблюдений:

Рассматриваем по числу переменных n вариантов выбора направления главных компонент с номером i (ГК $_i$). Для k -го варианта максимизируем вычисляемое отношение сигнал/шум у k -ой переменной $SNR_{ij}(k)$ — отношение среднего значения факторных нагрузок к среднеквадратичному отклонению этих величин для всех наблюдений:

$$SNR_{ij}(k) = \frac{m_{ij}(k)}{s_{ij}(k)},$$

где $m_{ij}(k) = \frac{1}{T} \sum_{t=1}^T w^t_{ij}(k)$ — средние значения нагрузок в наборе из T i -ых главных компонент для k -го варианта, $s_{ij}^2 = \frac{1}{T} \sum_{t=1}^T (w^t_{ij}(k) - m_{ij}(k))^2$ — среднеквадратичное отклонение значения нагрузок для k -го варианта; $i, j, k = 1, \dots, n$.

Максимальное значение вычисляемого отношения сигнал/шум у k -ой переменной $SNR_{ij}(k)$ достигается, если все нагрузки k -ой переменной имеют одинаковые знаки.

Нагрузки, для которых вычисленное значение отношения сигнал/шум $SNR_{ij}(k)$ не менее порогового значения

$$SNR_{ij}(k) \geq SNR_{thr.},$$

считаем *значимыми*, а переменные с такой нагрузкой *действующими*. Сумму вычисленных значений отношения сигнал/шум у действующих переменных для варианта k запоминаем:

$$S_{1i}(k) = \sum_j SNR_{ij}(k),$$

где j — номер действующей переменной для k -го варианта выбора направления ГК $_i$.

Из n рассмотренных вариантов выбора направлений выбирается вариант k^* , для которого сумма отношений сигнал/шум у действующих переменных максимальна

$$k^* : S_{1i}(k^*) = \max_k S_{1i}(k), \quad S_{1i} = S_{1i}(k^*).$$

- определение эмпирических главных компонент $w_i = \langle w_{i1}, w_{i2}, \dots, w_{in} \rangle$, $i = 1, \dots, n$: факторные нагрузки у *недействующих* переменных обнуляются, для *действующих* переменных в качестве значения факторной нагрузки принимаем

среднее значение по всем наблюдениям (при согласовании направлений собственных векторов):

$$w_{ij} = \frac{1}{T} \sum_{t=1}^T w_{ij}^t(k^*);$$

– выбор числа эмпирических главных компонент (ЭГК) N ($N \leq n$).

Количество выбранных ЭГК N должно обеспечивать максимально возможную информативность вычисляемой интегральной характеристики γ :

$$\gamma = \gamma_{\sigma} \cdot \gamma_{SNR} \quad ,$$

где $\gamma_{SNR} = \frac{S_{11} + S_{12} + \dots + S_{1N}}{S_{21} + S_{22} + \dots + S_{2N}}$ — SNR-информативность,

$\gamma_{\sigma} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_N}{\lambda_1 + \lambda_2 + \dots + \lambda_n}$ — дисперсионная информативность.

4 шаг. Согласование направлений эмпирических главных компонент

– нормирование ЭГК умножением на квадратный корень из соответствующего эмпирического собственного числа:

$$w^*_i = \langle \sqrt{\lambda_i} \cdot w_{i1}, \sqrt{\lambda_i} \cdot w_{i2}, \dots, \sqrt{\lambda_i} \cdot w_{in} \rangle = \langle w^*_{i1}, w^*_{i2}, \dots, w^*_{in} \rangle;$$

– согласование направлений нормированных ЭГК.

Критерием выбора направления ЭГК является максимизация значений целевой функции, вычисляемое как свертка значений данных на компоненты вектора весов. Максимальное значение неотрицательных переменных есть единица, поэтому для максимизации значений целевой функции выбираем из двух возможных направлений эмпирической главной компоненты то, при котором сумма ее координат максимальна;

– нахождение единого для всех наблюдений вектора весовых коэффициентов $W^* = \langle w^*_{11}, w^*_{12}, \dots, w^*_{1n} \rangle$ суммированием нагрузок согласованных эмпирических главных компонент $w^*_j = \sum_{i=1}^N w_{ij}^*$, $j = 1, \dots, n$, $i = 1, \dots, N$.

– вычисление интегрального показателя для наблюдения t как суммы произведений значений вектора весов и матрицы унифицированных показателей A^t

$$q^t = A^t \cdot W^* . \quad (5)$$

Для фиксированного момента t интегральную оценку для объекта с номером i записывают в виде аддитивной свертки данных с весами, определяемыми вектором W^* :

$$q_i^t = \sum_{j=1}^n w_j^* \cdot a_{ij}^t, \quad i = 1, \dots, m, \quad j = 1, \dots, n. \quad (6)$$

5 шаг. Определение весовых коэффициентов блоков

– вес блока l \tilde{w}_l , где $l = 1, \dots, L$, (L — количество блоков) пропорционален величине суммы действующих ОСШ этого блока S_1^l :

$$\tilde{w}_l = \frac{S_1^l}{S_1^1 + S_1^2 + \dots + S_1^L}, \quad \sum_{l=1}^L \tilde{w}_l = 1, \quad \tilde{w}_l \geq 0, \quad S_1^l = \sum_{i=1}^{N_l} S_{1i}^l,$$

где N_l — число выбранных эмпирических главных компонент в блоке l .

6 шаг. Вычисление интегральной характеристики системы

– вычисляем характеристику всей системы, учитывая веса блоков:

$$y_i^t = \sum_{l=1}^L \tilde{w}_l \cdot q_{il}^t,$$

где q_{il}^t — интегральная оценка блока l для объекта с номером i в момент t .

Шаги 2,3,4 и 5 выполняются независимо для каждого блока. На шаге 6 используются результаты предыдущих шагов.

4. Вычисление интегральных индикаторов качества жизни населения с помощью предложенного алгоритма

Рассмотрим работу алгоритма для вычисления интегральной характеристики качества жизни населения субъектов Российской Федерации. Выбор переменных и деление их на блоки является принципиальным вопросом, лежащим вне рассматриваемого обсуждения. Поэтому воспользуемся списком переменных из исследования, выполненного под руководством С.А. Айвазяна (табл.1), являющегося признанным авторитетом в исследовании качества жизни [7]. Деление переменных на блоки также произведено авторами исследования [7]. Отсутствующие в открытом доступе данные заменим на близкие к ним (выделено цветом в табл.1). Все значения переменных взяты из открытых справочников Росстата [12].

Вменение отсутствующих значений является уязвимым местом методик вычисления интегральной характеристики. Наличие ряда наблюдений во многом снимает эту проблему. Единичные отсутствующие значения восстановлены линейной интерполяцией по соседним имеющимся, в случае отсутствия значений показателей для ряда лет их дополняем их средними значениями по тому федеральному округу, к которому относится рассматриваемый субъект Федерации.

Среди перечисленных переменных переменные 1, 2, 5, 7, 9, 10, 12, 21, 22, 23 связаны с вычисляемой характеристикой монотонной возрастающей зависимостью, когда оптимальное значение j - го показателя максимально. Для остальных показателей, кроме переменной 27, оптимальное значение показателя минимально. Для переменной 27 «Коэффициент миграционного прироста» примем $x_j^{opt} = 0$, $j=27$. Значение соответствует ситуации, когда количество выбывших из региона совпадает с количеством приехавших.

Определение четвертой эмпирической главной компоненты представлено в табл.2. Выбранные направления собственных векторов максимизируют сумму значений ОСШ у действующих переменных, для которых отношение сигнал/шум — отношение среднего к среднеквадратичному отклонению — не менее 2,2. В этой ЭГК оказалась одна действующая переменная (выделена темным цветом в табл.2). Факторные нагрузки этой переменной в ЭГК определит средняя величина нагрузки по годам, остальные переменные в ЭГК обнуляются (так же, как в методах факторного анализа).

Таблица 1

Переменные для вычисления интегральных индикаторов качества жизни населения

Уровень благосостояния населения	
1	Отношение ВРП на душу населения к прожиточному мин, единиц
2	Покупательская способность среднедушевых денежных доходов по отношению к наборам прожиточного минимума (%)
3	Доля численности населения с денежными доходами ниже прожиточного минимума (%)
4	Отношение совокупных доходов 20% самых богатых и 20% самых бедных
5	Обеспеченность населения собственными легковыми автомобилями (автомобилей/1 000 чел.)
6	Доля семей, состоящих на учете на получение жилья (%)
7	Приходится общей площади жилищного фонда на одного жителя (кв. м/10 чел.)
8	Доля ветхого и аварийного жилья (%)
9	Плотность автомобильных дорог общего пользования (км/10 000 кв. км)
Качество населения	
10	Ожидаемая продолжительность жизни при рождении (лет)
11	Число умерших детей в возрасте до 1 года на 1000 населения (чел.)
12	Коэффициент естественного прироста (на 1000 населения)
13	Число умерших от инфекционных и паразитарных болезней и туберкулеза на 100 000 нас.
14	Число умерших от новообразований на 100 000 населения (чел.)
15	Число умерших от болезней системы кровообращения на 100 000 населения (чел.)
16	Число умерших от болезней органов дыхания на 100 000 населения (чел.)
17	Число умерших от болезней органов пищеварения на 100 000 населения (чел.)
18	Заболеваемость от травм, отравлений и некоторых других внешних причин на 100 000 нас.
19	Число инвалидов на 1 000 населения (чел.)
20	Зарегистрировано случаев заболевания врожденными аномалиями на 1 000 населения
21	Доля специалистов с высшим образованием среди занятых в экономике (%)
22	Приведенная производительность труда (ВРП/среднегодовая численность занятых в экономике, тыс. руб./чел.)
23	Количество выпускников высших и средних учебных заведений на 1000 населения
Качество социальной сферы	
24	Уровень безработицы (%)
25	Доля работников занятых во вредных и опасных условиях труда в среднегодовой численности занятых в экономике (%)
26	Численность пострадавших на производстве со смертельным исходом или с утратой трудоспособности на 1 рабочий день и более на 1 000 работающих
27	Коэффициент миграционного прироста на 10 000 населения
28	Число зарегистрированных умышленных убийств и покушений на убийство на 100 000 нас.
29	Число зарегистрир. фактов умышл. причинения тяжкого вреда здоровью на 100 000 нас.
30	Число зарегистрированных изнасилований и покушений на изнасилование на 100 000 нас.
31	Число разбоев, грабежей, краж из квартир граждан в расчете на 100 000 населения
32	Число зарегистрированных незаконных присвоений или растрат в расчете на 100 000 нас.
33	Число больных, состоящих на учете с диагнозом наркомания и токсикомания на 100 000 нас.
34	Число больных, состоящих на учете с диагнозом алкоголизм в расчете на 100 000 нас.
35	Число больных, инфицированных туберкулезом в расчете на 100 000 населения
36	Смертность от внешних причин на 100 000 населения
37	Число больных психическими расстройствами на 100 000 населения

Таблица 2

Определение четвертой эмпирической главной компоненты Блока 1

4 ГК	Переменные								
	1	2	3	4	5	6	7	8	9
2007	0,21	-0,23	-0,31	-0,09	0,42	0,46	-0,14	0,04	-0,62
2008	0,08	-0,07	-0,44	-0,35	0,37	0,32	-0,56	0,31	-0,16
2009	-0,06	0,04	-0,36	-0,43	0,18	0,39	-0,64	0,30	0,03
2010	-0,01	-0,04	-0,42	-0,40	0,31	0,18	-0,60	0,40	0,12
2011	-0,10	-0,02	-0,40	-0,41	0,27	-0,07	-0,47	0,60	0,07
2012	-0,03	-0,04	-0,37	-0,34	0,22	-0,30	-0,29	0,72	0,02
2013	0,02	0,00	-0,20	-0,05	-0,03	-0,69	0,24	0,63	-0,20
2014	-0,06	-0,04	-0,28	-0,26	-0,37	0,27	-0,29	0,72	-0,20
<i>Среднее, т</i>	0,01	-0,05	-0,35	-0,29	0,17	0,07	-0,34	0,46	-0,12
<i>Выб. ср.кв.др.откл., s</i>	0,10	0,08	0,08	0,15	0,26	0,40	0,29	0,24	0,24
ОСШ	0,07	0,62	4,26	1,97	0,67	0,18	1,17	1,92	0,49
	Сумма ОСШ по строке								11,34
	Действующие ОСШ								4,26

В табл.3 представлены все нормализованные и согласованные по направлению эмпирические главные компоненты (ЭГК) первого блока. Поскольку четвертое собственное число $\lambda_4 = 0,79 < 1$, значения факторной нагрузки в этой ЭГК уменьшились по сравнению с табл.2. Направления ЭГК выбираются, чтобы максимизировать значение вычисляемой линейной свертки определяемых весов (как правило, положительных) с переменными, чьи значения лежат на интервале $[0, 1]$.

Таблица 3

Эмпирические главные компоненты Блока 1

№ ГК	Эмпирические собственные числа	Переменные								
		1	2	3	4	5	6	7	8	9
1	3,34	0,80	0,90	0,82	-0,75	0,53		0,38	0,31	
2	2,04	-0,27	-0,31		0,43	0,43	0,64	0,64	0,64	
3	1,14		0,19			-0,37			0,18	0,79
4	0,79			0,31						
5	0,61								0,34	
6	0,50					0,43			-0,26	
7	0,31	0,35	-0,12	-0,24						0,19
8	0,24				0,26			-0,20	0,12	
9	0,03		0,14	-0,08	0,08					
Вектор весов		0,88	0,79	0,81	0,02	1,02	0,64	0,83	1,33	0,98

Максимальное значение, вносимое ЭГК в композитный индекс, достигается, если все значения переменных равны единице, т.е. определяется суммированием факторных нагрузок ЭГК. Если сумма факторных нагрузок у ЭГК отрицательна, значение направление меняем. Так четвертая ЭГК, определяемая табл.2, изменила направление.

Искомый вектор весов определяется суммированием одноименных переменных выбранного числа эмпирических главных компонент, представленных в табл.3. Число та-

ких компонент выбирается, чтобы максимизировать информативность полученного решения согласно (4).

Таблица 4

Определение информативности интегрального показателя Блока 1.

Уровень благосостояния населения

	Номер эмпирической главной компоненты								
	1	2	3	4	5	6	7	8	9
Сумма ОСШ k -ой ЭГК	102,2	57,6	19,7	11,3	7,9	10,7	20,9	16,5	100,5
Сумма действующих ОСШ k -ой ЭГК	101,4	55,0	15,4	4,2	3,4	4,9	16,5	10,8	95,7
Накопленный вклад, γ_{SNR}	0,99	0,98	0,96	0,92	0,90	0,88	0,87	0,86	0,88
Эмпирические собственные числа	3,34	2,04	1,14	0,79	0,61	0,50	0,31	0,24	0,03
Накопленный вклад, γ_{σ}	0,37	0,60	0,72	0,81	0,88	0,94	0,97	1,00	1,00
Информативность, $\gamma_{\sigma} \cdot \gamma_{SNR}$	0,37	0,58	0,69	0,75	0,79	0,82	0,85	0,85	0,88

В табл.4 приведен пример определения информативности интегрального показателя Блока 1 «Уровень благосостояния населения». При рассмотрении всех 9 ЭГК суммарная информативность максимальна и составит около 88%. Поэтому для вычисления интегрального показателя используем все эмпирические главные компоненты. Чем больше переменных описывают систему, тем меньшее их относительное количество участвует в построении композитного индекса. Для второго блока выбираем 11 из 14 ЭГК, для третьего — 10 из 14 ЭГК. (В другом случае в системе, которую описывают 51 переменных, было выбрано 21 ЭГК). Веса блоков определяем пропорционально действующим ОСШ переменных этого блока (табл.5), что аналогично пропорциональности силе принятого сигнала. Блок 3 «Качество социальной сферы», в котором 14 переменных, оказался менее значим, чем блок 1 «Уровень благосостояния населения», в котором 9 переменных.

Таблица 5

Определение весов блоков

Блок	1	2	3
Сумма действующих ОСШ	307,4	397,7	254,3
Вес блока	0,32	0,41	0,27

Основной недостаток методик определения весов с помощью аппарата факторного анализа состоит в необоснованных резких изменениях рейтингов объектов для последовательных наблюдений. При рассмотрении коротких временных промежутков при отсутствии катастроф и потрясений естественно ожидать плавное изменение как регистрируемых показателей, так и интегрального показателя качества системы и рейтингов, выставляемых на основании этого показателя. Отсутствие резких колебаний является естественным следствием предположения об адекватности рассматриваемой модели объективной реальности.

В проведенном исследовании число объектов, где значения рейтинга изменились по сравнению с предыдущим годом более чем на 15% от возможного максимального изменения, составило 1,1% случаев. Для сравнения, в работах [3, 4], где вычислялись интегральные характеристики муниципальных образований Тюменской и Самарской области для ряда лет, эта величина составляет 42,7% и 43,2%. Среднее изменение рейтинга за год в этих исследованиях составляет около 20%, в приведенном исследовании — 3,7%. Приведенную в [3, 4] интегральную характеристику нельзя назвать надежной, и на её основе оценивать качество жизни субъектов РФ. Источником неудовлетворительного качества вычисленных характеристик может являться, в частности, погрешность исходных данных, которую не учитывает классический метод главных компонент. Отметим, что ранги 25 развитых европейских стран за 2009–2011 годы, выставленные на основании значений ИРПЧ, дают среднее изменение рейтинга за год 7,7% [18].

На рис.2 показаны изменения интегральной характеристики качества жизни для Уральского ФО. Ожидаемо северные субъекты округа, ориентированные на добычу нефти и газа, имеют более высокие показатели качества жизни, чем промышленные области. За 100% принято максимальное значение интегральной характеристики в 2007 году (Москва), за 0% — минимальное значение в 2007 году (Тыва).

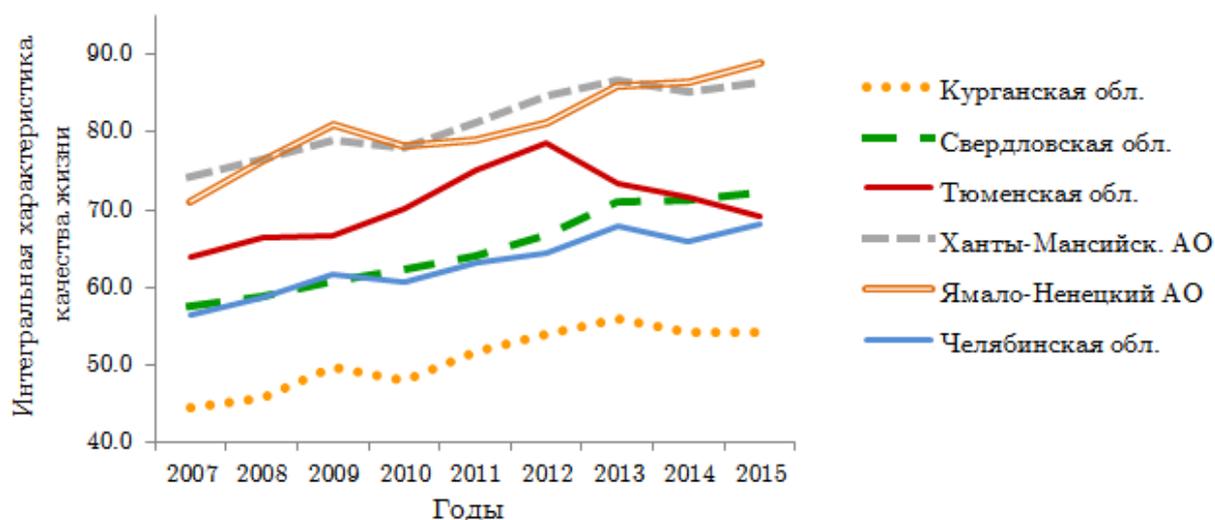


Рис. 2. Интегральная характеристика качества жизни некоторых субъектов Российской Федерации за 2007–2014 годы (2015 год — прогноз)

Заключение

В работе рассмотрено решение задачи построения латентной интегральной характеристики изменения качества системы на основании регистрируемых измерений для ряда наблюдений. Задача относится к классу задач выделения полезного сигнала из массива данных в условиях априорной неопределенности и решается на основании задаваемого отношения сигнал/шум. Предложен алгоритм построения интегральных характеристик с определением неслучайных составляющих главных компонент, характеризующих структуру рассматриваемой системы. Также предложен новый подход к выбору числа главных компонент, к определению весов рассматриваемых подсистем и к определению информативности полученной характеристики на основании дисперсионного критерия и выбранного параметра — отношения сигнал/шум. Алгоритм был использован для получения интегральных индикаторов качества жизни населения субъектов Российской Фе-

дерации. Предлагаемая методика может быть использована для вычисления интегральных оценок изменения качества плохо формализуемых систем. В дальнейшем следует проанализировать результаты применения методики построения интегральной характеристики качества системы не для заведомо хорошего (и не очень большого) набора входных переменных, а используя широкую базу измерений разных социальных систем, фиксируемую Росстатом.

Работа выполнена при финансовой поддержке проектной части государственного задания в сфере научной активности Министерства образования и науки Российской Федерации, проект № 1.949.2014/К.

Литература

1. Айвазян С.А. Интегральные индикаторы качества жизни населения: их построение и использование в социально-экономическом управлении межрегиональных сопоставлениях. М.: ЦЭМИ РАН, 2000. 56 с.
2. Айвазян С.А. К методологии измерения синтетических категорий качества жизни населения // Экономика и математические методы. 2003. Т. 39. № 2. С. 33–53.
3. Айвазян С.А., Степанов В.С., Козлова М.И. Измерение синтетических категорий качества жизни населения региона и выявление ключевых направлений совершенствования социально-экономической политики (на примере Самарской области и ее муниципальных образований) // Прикладная Эконометрика. 2009. №3(19). С. 18–84.
4. Гайдамак И.В., Хохлов А.Г. Моделирование интегральных показателей качества жизни населения юга Тюменской области // Вестник Тюменского государственного университета. 2009. № 6. С. 176–186.
5. Голяндина Н.Э., Усевич К.Д., Флоринский И.В. Анализ сингулярного спектра для фильтрации цифровых моделей // Геодезия и картография. 2008. №5. С. 21–28.
6. Голяндина Н.Э. Метод «Гусеница» — SSA: анализ временных рядов. СПб.: Санкт-Петербургский университет, 2004. 74 с.
7. Исакин М.А. Модификация метода k -средних с неизвестным числом классов // Прикладная эконометрика. 2006. Выпуск № 4. С. 62–70.
8. Жгун Т.В. Построение интегральной характеристики изменения качества системы на основании статистических данных как решение задачи выделения сигнала в условиях априорной неопределенности // Вестник Новгородского государственного университета. Серия: Технические науки. 2014. № 81. С. 10–16.
9. Жгун Т.В. Построения интегральной характеристики демографического развития территорий на примере муниципальных образований Новгородской области // Региональная экономика: теория и практика. 2013. № 36(315), сентябрь. С. 2–12.
10. Жгун Т.В. Вычисление интегрального показателя эффективности функционирования динамической системы на примере интегральной оценки демографического развития муниципальных образований Новгородской области // Вестник Новгородского государственного университета. Серия: Физико-математические науки. 2013. №75. Т.2. С.11–16.
11. Жгун Т.В. Исследование формальных методов построения латентной характеристики качества систем // Вестник Новгородского государственного университета. Серия: Физико-математические науки. 2014. № 80. С. 13–19.

12. Федеральная служба гос. статистики / Федеральная служба гос. статистики. — М., 2002–2014. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/stats/publications/catalog/doc_1137674209312(дата обращения: 13.01.2016).
13. Bandura R., del Campo C.M. Indices of National Performance: A Survey. Office of Development Studies, United Nations Development Programme, New York, November 17, 2006. URL: http://www.eldis.org/vfile/upload/1/document/1112/measuring_country_performance_2006update.pdf (дата обращения 01.07.2016).
14. Bandura R. A Survey of Composite Indices Measuring Country Performance: 2008 Update. Office of Development Studies United Nations Development Programme, New York, 2008/ 96 p. URL: http://web.undp.org/developmentstudies/docs/indices_2008_bandura.pdf (дата обращения 01.07.2016).
15. Bandura R. Composite Indicators and Rankings: Inventory 2011. Working Paper. Abbreviated version. 257 p. URL: <http://nebula.wsimg.com/beb0f5b0e9b6f0c298ae72e7519b22a4> (Дата обращения 01.07.2016).
16. Foa R.; Tanner J.C. Methodology of the Indices of Social Development / ISD Working Paper Series from International Institute of Social Studies of Erasmus University Rotterdam (ISS), The Hague. 2012. No 4. .66 p. URL: <http://repub.eur.nl/pub/50510/ISD-WP-2012-4.pdf> (дата обращения 01.07.2016).
17. Handbook on Constructing Composite Indicators: Methodology and User Guide. – OECD Publication. Paris CEDEX 16. 2008. 162 p.
18. Human Development Reports. 1990–2013. /United Nations Development Programme, 1990–2014. URL: <http://hdr.undp.org/en/reports/> (дата обращения: 13.06.2016).
19. Hightower W.L. Development of an index of health utilizing factor analysis // Medical Care, 1978. Vol. 16. P. 245–55.
20. Krishnan V. Constructing an Area-based Socioeconomic Index: A Principal Components Analysis Approach. Early Child Development Mapping Project (ЕСMap), Community–University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton, Alberta T5J 4P6, Canada. URL: http://www.cup.ualberta.ca/wp-content/uploads/2013/04/SEICUPWebsite_10April13.pdf (дата обращения 01.07.2016).
21. Lindman C., Sellin J. Measuring Human Development. The Use of Principal Component Analysis in Creating an Environmental Index. University essay from Uppsala universitet, Uppsala, 2011. 45 p. URL: <http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A464378&dswid=-5179> (дата обращения 01.07.2016).
22. McKenzie, D.J. Measuring Inequality with Asset Indicators. Journal of Population Economics. 2005. Vol. 18, Issue 2. P. 229–260.
23. Nardo M., Saisana M., Saltelli A., Tarantola S. Tools for composite indicators building. european commission, EUR 21682 EN. Joint Research Centre, Ispra, Italy.2005/URL: <http://farmweb.jrc.ec.eu.int/ci/bibliography.htm> (дата обращения 01.07.2016).
24. Nicoletti G., Scarpetta S., Boylaud O. Summary indicators of product market regulation with an extension to employment protection legislation, Economics department working papers NO. 226, ECO/WKP(99)18. 2000. URL: <http://www.oecd.org/eco/eco> (дата обращения 01.07.2016).
25. Saltelli A. Composite indicators between analysis and advocacy // Social Indicators Research, March 2007. Vol. 81, Issue 1. P. 65–77

26. Saltelli A., Munda G., Nardo M. From Complexity to Multidimensionality: the Role of Composite Indicators for Advocacy of EU Reform // Tijdschrift voor Economie en Management. 2006. Vol. LI, No 3 URL: http://feb.kuleuven.be/rebel/jaargangen/2001-2010/2006/TEM%202006-3/TEM_2006-3_03_Saltelli.pdf (дата обращения 01.07.2016).
27. Sharpe A. Literature Review of Frameworks for Macro-indicators, Ottawa, Canada: Centre for the Study of Living Standards. 2004. URL: <https://ideas.repec.org/p/sls/resrep/0403.html> (дата обращения 01.07.2016).
28. Somarriba N., Pena B. Synthetic Indicators of Quality of Life in Europe // Social Indicators Research, Oct. 2009. Vol. 94, No 1. P. 115-133.
29. Tarantola S., Saisana M., Saltelli A. Internal Market Index 2002: Technical details of the methodology. JRC European Commission. Institute for the Protection and Security of the Citizen Technological and Economic Risk Management Unit I-21020 Ispra (VA) Italy. 2002. URL: http://ec.europa.eu/internal_market/score/docs/score11/im-index-2002_en.pdf (дата обращения 01.07.2016).
30. Vyas S., Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Published by Oxford University Press in association with The London School of Hygiene and Tropical Medicine. All rights reserved. DOI:10.1093/heapol/czl029 Advance Access publication 9 October 2006. URL: <http://heapol.oxfordjournals.org> (дата обращения 01.07.2016).

Жгун Татьяна Валентиновна, к.ф.-м.н., доцент, кафедра прикладной математики и информатики, Новгородский государственный университет имени Ярослава Мудрого (Великий Новгород, Российская Федерация)

DOI: 10.14529/cmse170101

AN ALGORITHM FOR CONSTRUCTING INTEGRAL QUALITY INDICATOR OF COMPLEX SYSTEMS FOR A SEQUENCE OF OBSERVATIONS

© 2017 T. V. Zhgun

Novgorod State University a. Yaroslav the Wise

(Bolshaya Sankt-Peterburgskaya Avenue, 41, Veliky Novgorod, 173003 Russia)

E-mail: Tatyana.Zhgun@novsu.ru

Received: 26.10.2016

In the paper we propose an algorithm of constructing the integral characteristics of changes in the quality system based on the recorded measurements. The algorithm provides allotment signal in a multidimensional array of data in terms of a priori uncertainty on the basis of the specified signal-to-noise ratio. Constructing of latent integral characteristics of changes of the quality system on the basis of statistical indicators for a number of consecutive observations is based on the principal component method, taking into account the presence of noise in the measured data (SNR-based algorithm). In classical PCA informativity of integral characteristics is given a priori and is provided by selecting the number of principal components. In the proposed algorithm the information content of the solution is evaluated a posteriori on the basis of variance criterion and the selected parameter signal-to-noise ratio. Based on the proposed algorithm we construct integral indicators of the quality of life of subjects of the Russian Federation for the years of 2007-2014.

Keywords: quality management system, an integral characteristic of the quality, principal component analysis change the quality characteristics, the integral indicators of quality of life, noise in the measured data, the signal-to-noise ratio, informativity of the principal component analysis.

FOR CITATION

Zhgung T.V. An algorithm for constructing integral quality indicator of complex systems for a sequence of observations. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2017. vol. 6, no. 1. pp. 5–25. (in Russian) DOI: 10.14529/cmse170101.

References

1. Ayvazyan S.A. Integral'nye indikatory kachestva zhizni naselenija: ih postroenie i ispol'zovanie v social'no-jekonomicheskom upravlenii mezhregional'nyh sopostavlenijah. [Integrated Indicators of Quality of Life of the Population: Their Construction and use in Social and Economic Management Interregional Comparisons]. M.: TsEMI Russian Academy of Sciences, 2000. 56 p.
2. Ayvazyan S.A. *K metodologii izmerenija sinteticheskikh kategorij kachestva zhizni nasele-nija*. [To Methodology of Measurement of Synthetic Categories of Quality of Life of the Population]. *Jekonomika i matematicheskie metody*. [Economy and Mathematical Methods]. 2003. vol. 39, no. 2. pp. 33–53. (in Russian)
3. Ayvazyan S.A., Stepanov V.S., Kozlova M.I. *Izmerenie sinteticheskikh kategorij kachestva zhizni naseleniya regiona i vyyavlenie klyuchevykh napravleniy sovershenstvovaniya sotsial'no-jekonomicheskoy politiki (na primere Samarskoy oblasti i ee munitsipal'nykh obrazovaniy)* [Measurement of Synthetic Categories of the Quality of Life of Population in the Region and Identify Key Areas of Improvement in Socio-Economic Policy (on the Example of Samara Region and its Municipalities)]. *Prikladnaja Jekonometrika* [Applied Econometrics]. 2009. no. 3(19). pp. 18–84. (in Russian)
4. Gaydamak I.V., Khokhlov A.G. *Modelirovanie integral'nykh pokazateley kachestva zhizni naseleniya yuga Tyumenskoy oblasti* [Modeling of Integral Indicators of the Quality of Life of the Population of the South of the Tyumen Region]. *Vestnik Tjumenskogo gosudarstvennogo universiteta*. [Bulletin of the Tyumen State University]. 2009. no. 6. pp. 176–186. (in Russian)
5. Goljandina N.J., Usevich K.D., Florinskij I.V. *Analiz singular'nogo spektra dlja fil'tracii cifrovyyh modelej* [The Analysis of a Singular Range for a Filtration of Digital Models]. *Geodeziya i kartografiya*. [Geodesy and Cartography] 2008, no. 5. pp. 21–28. URL: <http://iflorinsky.narod.ru/Florinsky-2008ts.pdf>. (accessed 13.01.2015). (in Russian)
6. Goljandina N.J. *Metod "Gusenica" — SSA: analiz vremennykh rjadov* [The Method of "Caterpillar" — SSA: Analysis of Time Series, St. Petersburg.: St. Petersburg University Press, 2004. 74 p.
7. Isakin M.A. *Modifikacija metoda k –srednih s neizvestnym chislom klassov* [Modification of a Method to k -Means with Unknown Number of Classes]. *Prikladnaya ekonometrika* [Applied Econometrics]. 2006. Release no. 4. pp. 62–70. (in Russian)
8. Zhgung T.V. *Postroenie integral'noj harakteristiki izmenenija kachestva sistemy na osnovanii statisticheskikh dannykh kak reshenie zadachi vydelenija signala v uslovijah apriornoj neopredelennosti* [Creation of the Integrated Characteristic of Change of Quality of System on the Basis of Statistical Data as the Solution of a Problem of Allocation of a Signal in the Conditions of Aprioristic Uncertainty]. *Vestnik Novgorodskogo gosudarstvennogo universiteta. Seriya: Tekhnicheskie nauki* [Bulletin of Novgorod State University. Series: Engineering sciences.]. 2014. no. 81. pp. 10–16. (in Russian)

9. Zhgun T.V. *Postroeniya integral'noj harakteristiki demograficheskogo razvitija territorij na primere municipal'nyh obrazovaniy Novgorodskoj oblasti* [Creation of the Integrated Characteristic of Demographic Development of Territories on the Example of Municipalities of the Novgorod Region] *Regional'naya ekonomika: teoriya i praktika* [Regional Economy: Theory and Practice]. 2013. no. 36(315), September. pp. 2–12. (in Russian)
10. Zhgun T.V. *Vychislenie integral'nogo pokazatelya jeffektivnosti funkcionirovaniya dinamicheskoy sistemy na primere integral'noj ochenki demograficheskogo razvitija municipal'nyh obrazovaniy Novgorodskoj oblasti* [Calculation of an Integrated Indicator of Efficiency of Functioning of Dynamic System on the Example of an Integrated Assessment of Demographic Development of Municipalities of the Novgorod Region]. *Vestnik Novgorodskogo gosudarstvennogo universiteta. Seriya: Fiziko–matematicheskie nauki* [Bulletin of Novgorod State University. Series: Physico–Mathematical Sciences]. 2013. no. 75. vol. 2. pp. 11–16. (in Russian)
11. Zhgun T.V. *Issledovanie formal'nyh metodov postroeniya latentnoj harakteristiki kachestva system* [Research of Formal Methods of Creation of the Latent Characteristic of Quality of Systems. *Vestnik Novgorodskogo gosudarstvennogo universiteta Seriya: Fiziko–matematicheskie nauki* [Bulletin of Novgorod State University. Series: Physico–Mathematical Sciences]. 2014. no. 80. pp. 13–19. (in Russian)
12. *Federal'naja sluzhba gos. Statistiki* [Federal service state. statistics]. Federal Service State. Statistics.–M, 2002–2014. URL: http://shchshchshch.gks.ru/shchps/shchtsm/tsonetst/rosstat_main/rosstat/ru/statistitss/publitsations/tsatalog/dots_1137674209312 (accessed 13.01.2016).
13. Bandura R., del Campo C.M. *Indices of National Performance: A Survey*. Office of Development Studies, United Nations Development Programme, New York, November 17, 2006. 91 p. URL: http://www.eldis.org/vfile/upload/1/document/1112/measuring_country_performance_2006update.pdf (accessed: 01.07.2016).
14. Bandura R. *A Survey of Composite Indices Measuring Country Performance: 2008 Update*. Office of Development Studies United Nations Development Programme, New York, 2008/ 96 p. URL: http://web.undp.org/developmentstudies/docs/indices_2008_bandura.pdf (accessed: 01.07.2016).
15. Bandura R. *Composite Indicators and Rankings: Inventory*. 2011. Working Paper. Abbreviated version. 257 p. URL: <http://nebula.wsimg.com/beb0f5b0e9b6f0c298ae72e7519> (accessed: 01.07.2016).
16. Foa R., Tanner J. C. *Methodology of the Indices of Social Development / ISD Working Paper Series from International Institute of Social Studies of Erasmus University Rotterdam (ISS), The Hague*. 2012. No 4. P.66.
17. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publication. Paris CEDEX 16. 2008. 162 p.
18. *Human Development Reports. 1990–2013/ United Nations Development to the Program, 1990–2014*. URL: <http://hdr.undp.org/en/reports> (accessed 13.06.2016).
19. Hightower W. L. (1978). *Development of an Index of Health Utilizing Factor Analysis*. *Medical Care*. 1978. no. 16. pp. 245–55.
20. Krishnan V. *Constructing an Area–based Socioeconomic Index: A Principal Components Analysis Approach*. Early Child Development Mapping Project (ECMap), Community–University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton,

- Alberta T5J 4P6, Canada. URL: http://www.cup.ualberta.ca/wp-content/uploads/2013/04/SEICUPWebsite_10April13.pdf (accessed: 01.07.2016).
21. Lindman C., Sellin J. Measuring Human Development. The Use of Principal Component Analysis in Creating an Environmental Index. University essay from Uppsala universitet, Uppsala, 2011, 45 p. URL: <http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A464378&dsid=-5179> (accessed: 01.07.2016).
 22. McKenzie D. J. Measuring Inequality with Asset Indicators. *Journal of Population Economics*. 2005. vol. 18, issue 2. pp. 229–260.
 23. Nardo M., Saisana M., Saltelli A., Tarantola S. Tools for Composite Indicators Building. European Commission, EUR 21682 EN. Joint Research Centre, Ispra, Italy. 2005. URL: <http://farmweb.jrc.cec.eu.int/ci/bibliography.htm> (accessed: 01.07.2016).
 24. Nicoletti G., Scarpetta S., Boylaud O. Summary Indicators of Product Market regulation with an Extension to Employment Protection Legislation, Economics Department Working Papers NO. 226, ECO/WKP(99)18. 2000. URL: <http://www.oecd.org/eco/eco> (accessed: 01.07.2016).
 25. Saltelli A. Composite Indicators Between Analysis and Advocacy. *Social Indicators Research*. March 2007. vol. 81, issue 1. pp 65–77
 26. Saltelli A., Munda G., Nardo M. From Complexity to Multidimensionality: the Role of Composite Indicators for Advocacy of EU Reform. *Tijdschrift voor Economie en Management*. 2006. Vol. LI, No 3 URL: http://feb.kuleuven.be/rebel/jaargangen/2001-2010/2006/TEM%202006-3/TEM_2006-3_03_Saltelli.pdf (accessed: 01.07.2016).
 27. Sharpe A. Literature Review of Frameworks for Macro-indicators, Ottawa, Canada: Centre for the Study of Living Standards. 2004. URL: <https://ideas.repec.org/p/sls/resrep/0403.html> (accessed: 01.07.2016).
 28. Somarriba N., Pena B. Synthetic Indicators of Quality of Life in Europe. *Social Indicators Research*, 2009. vol. 94, issue 1. pp. 115–133.
 29. Tarantola S., Saisana M., Saltelli A. Internal Market Index 2002: Technical Details of the Methodology. JRC European Commission. Institute for the Protection and Security of the Citizen Technological and Economic Risk Management Unit I-21020 Ispra (VA) Italy. 2002. URL: http://ec.europa.eu/internal_market/score/docs/score11/im-index-2002_en.pdf (accessed: 01.07.2016).
 30. Vyas S., Kumaranayake L. Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis. Published by Oxford University Press in association with The London School of Hygiene and Tropical Medicine. DOI:10.1093/heapol/czl029 Advance Access publication 9 October 2006. URL: <http://heapol.oxfordjournals.org> (accessed: 01.07.2016).

МОДИФИКАЦИЯ МЕТОДА МАТРИЧНЫХ ПУЧКОВ, ИСПОЛЬЗУЮЩАЯ СОВМЕСТНОЕ ОЦЕНИВАНИЕ ПОЛЮСОВ СИГНАЛА И ОБРАТНЫХ К НИМ

© 2017 г. О.Л. Ибряева, Д.Д. Салов

Южно-Уральский государственный университет
(454080 Челябинск, пр. им. В.И. Ленина, д. 76),
E-mail: ibriaevaol@susu.ru, salovdd@yandex.ru

Поступила в редакцию: 20.10.2016

В статье рассматривается метод матричных пучков, являющийся параметрическим Прони-подобным методом обработки сигнала и позволяющим найти частоты, коэффициенты затухания, фазы и амплитуды суммы затухающих синусоид. Его преимуществом по сравнению с методом Прони является меньшее количество операций, что приводит к меньшей вычислительной ошибке. Предложена модификация данного метода способная решить задачу разделения истинных и ложных полюсов сигнала. Из отсчетов сигнала конструируются два пучка матриц, собственные значения которых в случае отсутствия шума совпадают с полюсами сигнала и обратными к ним. В случае зашумленного сигнала разделение истинных и ложных полюсов проводится за счет: 1) сингулярного разложения, 2) завышенного порядка предсказания, 3) анализа собственных значений двух пучков матриц. Приведен алгоритм модифицированного метода матричных пучков и сравнение этой модификации с классическим вариантом на модельном примере обнаружения сигнала в шуме. Показано, что классический метод не способен определить время начала полезного сигнала, поскольку подстраивает под шум сумму экспонент. Модифицированный метод матричных пучков решает задачу обнаружения сигнала в шуме и его параметров, т.к. помимо определения времени прихода сигнала, также способен оценить количество затухающих синусоид в нем и их параметры. Предложенный алгоритм обнаружения сигнала пригоден для работы с сигналами достаточно общего вида (суммой затухающих синусоид) и не требует знания законов распределения самого сигнала и его шумовой составляющей, как метод максимального правдоподобия.

Ключевые слова: обработка сигнала, метод матричных пучков, обнаружение сигнала в шуме, определения времени прихода, оценка частоты.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Ибряева О.Л., Салов Д.Д. Модификация метода матричных пучков, использующая совместное оценивание полюсов сигнала и обратных к ним // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. Т. 6, № 1. С. 26–37. DOI: 10.14529/cmse170102.

Введение

Метод матричных пучков (ММП) [1] является одним из Прони-подобных [2] методов оценки параметров суммы комплексных экспонент. Его преимуществом по сравнению с методом Прони является меньшее количество операций, что приводит к меньшей вычислительной ошибке. ММП находит многочисленные применения — в спектральном анализе сигналов ядерного магнитного резонанса [3], в идентификации радиолокационных объектов [4], в сейсмологии [5], медицине [6] и др.

Хорошо известно, что при использовании метода Прони в случае высокого уровня шума весьма затруднительно разделить корни характеристического полинома на корни,

соответствующие полезному сигналу и корни, соответствующие шуму. Для упрощения этого разделения и улучшения точности оценок полюсов сигнала используются три метода [7]: SVD-разложение, завышенный порядок предсказания и анализ нулей полиномов линейного предсказания вперед и назад, дающих оценки для полюсов сигнала Z и обратных к ним z^{-1} , соответственно. Нули, соответствующие истинным экспоненциальным сигналам, будут появляться во взаимно обратных точках, расположенных вдоль некоторого общего радиуса, что и облегчает идентификацию экспонент в шуме [8].

ММП, как и метод Прони, подстраивает сумму экспонент под любой аддитивный шум, присутствующий в данных и, по этой причине, не обеспечивает удовлетворительных результатов при значительном уровне шума. В исходном варианте ММП оценки для полюсов Z сигнала дают собственные значения некоего пучка матриц, сконструированного из отсчетов сигнала. Между тем, как показано в настоящей статье, несложно получить с помощью тех же данных оценку для z^{-1} и, на основе анализа значений Z и z^{-1} , решить вопрос разделения истинных экспоненциальных сигналов и шума. *Целью* настоящей работы является модификация метода матричных пучков, использующая совместное оценивание полюсов сигнала и обратных к ним и аналогичная описанной выше модификации метода Прони.

В качестве приложения данной модификации ММП в данной работе рассматривается задача определения параметров и времени прихода для сигнала с неизвестными параметрами. Насколько нам известно, ранее нигде ММП не использовался для определения времени прихода сигнала, возможно потому, что (как показано в данной статье) классический ММП плохо справляется с этой задачей. Заметим, что существует множество алгоритмов решения задачи обнаружения сигнала, но они либо работают преимущественно с синусоидальными сигналами (преобразование Фурье, обобщенный «кросс-коррелятор» [9–11]), либо требуют знания законов распределения как самого сигнала, так и его шумовой составляющей (метод максимального правдоподобия [12;13]). Предложенный в этой статье метод не требует такой априорной информации и работает с сигналами достаточно общего вида — суммой затухающих синусоид. Количество синусоид не предполагается известным заранее.

Структура работы такова. В разделе 1 мы описываем исходный вариант ММП. Его модификация и проблемы ее программной реализации обсуждаются в разделе 2. В разделе 3 приведен результат численного эксперимента с помощью модифицированного ММП. Рассматривается важная задача обнаружения сигнала (и оценки его параметров) в шуме. Показано, что, в отличие от классического ММП, предложенный в данной статье метод позволяет определять неизвестное время прихода сигнала и оценить его частоту. Завершается работа заключением с указанием направления будущих исследований.

1. Метод матричных пучков

ММП предназначен для нахождения параметров R_k , z_k сигнала вида:

$$x_n = \sum_{k=1}^M R_k z_k^n, \quad (1)$$

где $R_k = A_k e^{i\varphi_k}$ — комплексные амплитуды, $z_k = e^{(\alpha_k + i2\pi f_k)T}$ — комплексные экспоненты (полюсы сигнала), T — период дискретизации сигнала, $n = 0, 1, \dots, N-1$, N — число отсчетов сигнала.

Как правило, обрабатываемый сигнал является зашумленным, т.е. имеет вид:

$$y_n = x_n + v_n = \sum_{k=1}^M R_k z_k^n + v_n, \quad (2)$$

где v_n — отсчеты шума.

Из данных N отсчетов сигнала сформируем матрицы размеров $(N-L) \times L$:

$$Y_0 = \begin{pmatrix} y_{L-1} & y_{L-2} & \dots & y_0 \\ y_L & y_{L-1} & \dots & y_1 \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-2} & y_{N-3} & \dots & y_{N-L-1} \end{pmatrix}, \quad Y_1 = \begin{pmatrix} y_L & y_{L-1} & \dots & y_1 \\ y_{L+1} & y_L & \dots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-1} & y_{N-2} & \dots & y_{N-L} \end{pmatrix}. \quad (3)$$

Здесь L — единственный параметр метода матричных пучков, значение которого должно удовлетворять неравенству: $M \leq L \leq N - M$, где N — число отсчетов, M — число экспонент сигнала. Показано [14], что при выборе $\frac{N}{3} \leq L \leq \frac{N}{2}$ дисперсия оценки полюсов z_k будет минимальна, т.е. ММП будет наименее чувствителен к шуму.

Напомним [15], что обобщенная постановка задачи отыскания собственных значений выглядит следующим образом: Найти собственные векторы x и собственные значения λ , определяемые соотношением

$$Ax = \lambda Bx. \quad (4)$$

Характеристическое уравнение для обобщенной задачи о собственных значениях имеет вид: $\det(A - \lambda B) = 0$. Выражение $A - \lambda B$ называют *пучком матриц* и говорят о собственных значениях и собственных векторах пучка матриц.

Доказано [1], что собственными числами пучка матриц $Y_0 - \lambda Y_1$ в случае отсутствия шума являются полюсы сигнала $z_k = e^{(\alpha_k + i2\pi f_k)T}$. Данный факт и определил название метода.

Таким образом, найти z_k в случае чистого сигнала можно как собственные значения матрицы $Y_0^+ Y_1$. Однако в случае зашумленных данных собственные значения матрицы $Y_0^+ Y_1$ не будут в точности равны полюсам сигнала. Для эффективной фильтрации шума матрицу Y_0 необходимо предварительно подвергнуть операции сингулярного (SVD) разложения. Здесь и далее верхним индексом $+$ мы обозначаем псевдообратную матрицу Мура—Пенроуза. Напомним, что псевдообратная, в отличие от обратной, существует для любой матрицы, а операцию псевдообращения можно понимать [16] как решение задачи наилучшей аппроксимации (по методу наименьших квадратов с предельным вариантом регуляризации) для соответствующей системы уравнений.

Рассмотрим SVD-разложение матрицы Y_0 :

$$Y_0 = USV^H, \quad (5)$$

где U, V — унитарные матрицы размеров $(N-L) \times (N-L)$ и $L \times L$ соответственно, S — матрица размеров $(N-L) \times L$, на главной диагонали которой стоят неотрицательные

вещественные (сингулярные) числа в порядке невозрастания. Здесь и далее верхним индексом H мы обозначаем операцию эрмитова сопряжения.

Заметим, что в случае отсутствия шума диагональная матрица S имеет ровно M ненулевых сингулярных чисел, все последующие равны нулю. В случае зашумленного сигнала ненулевых сингулярных чисел уже не будет, однако между первыми M и последующими сингулярными числами матрицы S будет наблюдаться ярко выраженный скачок, который и позволит определить число комплексных экспонент в сигнале.

Итак, сингулярное разложение матрицы Y_0 позволяет определить число истинных экспонент сигнала M . Кроме того, оно может быть использовано для нахождения псевдообратной матрицы $Y_0^+ = VS^+U^H$. На практике используется усеченная до ранга M псевдообратная матрица $Y_0^+ = \sum_{i=1}^M \frac{1}{\sigma_{0i}} v_{0i} u_{0i}^H = V_0 A_0^{-1} U_0^H$, где $\sigma_{0i}, i = 1, \dots, M$ – наибольшие сингулярные числа матрицы Y_0 , v_{0i}, u_{0i} – соответствующие сингулярные векторы, $A_0 = \text{diag}(\sigma_{01}, \dots, \sigma_{0M})$.

После нахождения матрицы Y_0^+ для оценки полюсов сигнала $z_k, k = 1, \dots, M$, остается только найти M собственных чисел матрицы $Y_0^+ Y_1$ или, в силу следующей цепочки равенств:

$$Y_0^+ Y_1 q_k = z_k q_k, \quad (5)$$

$$V_0 A_0^{-1} U_0^H Y_1 q_k = z_k q_k, \quad (6)$$

$$A_0^{-1} U_0^H Y_1 V_0 (V_0^H q_k) = z_k (V_0^H q_k), \quad (7)$$

найти M собственных чисел матрицы $Z_E = A_0^{-1} U_0^H Y_1 V_0$, дающих оценку для полюсов сигнала $z_k = e^{(\alpha_k + i2\pi f_k)T}, k = 1, \dots, M$.

Комплексные амплитуды $R_k = A_k e^{i\phi_k}, k = 1, \dots, M$, находим, решая следующую задачу методом наименьших квадратов:

$$\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_M \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-1} & z_2^{N-1} & \dots & z_M^{N-1} \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_M \end{pmatrix}. \quad (8)$$

Описанный «классический» метод матричных пучков имеет следующий недостаток. При обработке реальных сигналов, особенно с большим уровнем шума, заметить скачок между сингулярными числами, соответствующими сигналу и шуму и провести тем самым разделение полюсов на истинные и ложные, оказывается весьма затруднительно. В методе Прони с этой проблемой справляется завышенный порядок предсказания: поскольку число истинных полюсов сигнала M не поддается определению, его берут явно завышенным, «с запасом». Далее для идентификации истинных полюсов среди всех найденных применяется анализ оценок для z_k и z_k^{-1} , даваемых нулями полиномов линейного предсказания вперед и назад.

В следующем разделе мы предлагаем аналогичную модификацию ММП.

2. Модификация метода матричных пучков

Как уже было отмечено в разделе 1, полюсы сигнала $z_k = e^{(\alpha_k + i2\pi f_k)T}$ являются собственными числами пучка матриц $Y_0 - \lambda Y_1$ и их можно найти как собственные числа матрицы $Y_0^+ Y_1$ или Z_E . Рассмотрим пучок матриц $Y_1 - \lambda Y_0$. Его собственными числами, очевидно, являются z_k^{-1} , которые, в отсутствие шума, можно найти как собственные числа матрицы $Y_1^+ Y_0$. В случае зашумленных данных следует рассмотреть матрицу $Y_1^+ Y_0$, где Y_1^+ — усеченная до ранга M псевдообратная матрица

$$Y_1^+ = \sum_{i=1}^M \frac{1}{\sigma_{1i}} v_{1i} u_{1i}^H = V_1 A_1^{-1} U_1^H. \quad (9)$$

Здесь $\sigma_{1i}, i=1, \dots, M$ — наибольшие сингулярные числа матрицы Y_1 , v_{1i}, u_{1i} — соответствующие сингулярные векторы, $A_1 = \text{diag}(\sigma_{11}, \dots, \sigma_{1M})$. Очевидно, что, как и в предыдущем разделе, можно вместо матрицы $Y_1^+ Y_0$ рассмотреть матрицу $Z_I = A_1^{-1} U_1^H Y_0 V_1$, т.к. их собственные числа совпадают.

Итак, в ММП, как и в методе Прони, можно получить оценки для полюсов z_k и обратных к ним z_k^{-1} , а значит можно предложить аналогичный способ отделения истинных полюсов от ложных, основанный на совместном анализе этих оценок. Из-за шума скачок между сингулярными числами будет размыт и число истинных полюсов сигнала M не будет найдено. Как и в методе Прони, используем завышенный порядок предсказания, выбрав M «с запасом».

Инициализация:

Задаем малое положительное число $\varepsilon = 0,01 \div 0,1$ и формируем матрицы Y_0, Y_1 из отсчетов сигнала;

Шаг 1. Находим усеченное SVD-разложение матриц

$$Y_0 = U_0 A V_0^H, \quad Y_1 = U_1 A_1 V_1^H$$

и предварительную оценку числа M полюсов сигнала.

Шаг 2. Составляем матрицы

$$Z_E = A^{-1} U_0^H Y_1 V_0, \quad Z_I = A_1^{-1} U_1^H Y_0 V_1.$$

Шаг 3. Находим их собственные числа $p_k, q_m, k, m = 1, \dots, M$.

Шаг 4. Определяем число M истинных полюсов сигнала и сами полюсы:

Если

$$\left| p_k - \frac{1}{q_m} \right| \leq \varepsilon \text{ для некоторых } k, m,$$

То p_k — истинный полюс сигнала,

Иначе p_k — ложный полюс.

Шаг 5. Находим комплексные амплитуды $R_k = A_k e^{i\varphi_k}, k = 1, \dots, M$.

Стоп.

Рис. 1. Модифицированный метод матричных пучков

Далее для отделения истинных полюсов сигнала от полюсов, соответствующих шуму, проанализируем собственные значения матриц Z_E , Z_I . Из них ровно M собственных значений (соответствующие z_k и z_k^{-1}) будут находиться во взаимно обратных точках, что позволяет идентифицировать истинные полюсы.

На рис. 1 представлен алгоритм модифицированного метода матричных пучков.

Замечание 1. Комплексные амплитуды, так же, как и в «классическом» ММП, находятся из решения задачи (8).

Замечание 2. При программной реализации приведенного алгоритма из-за шума возникает проблема несовпадения собственных чисел p_k , q_m матриц Z_E , Z_I с числами z_k , z_k^{-1} . По этой причине числа p_k , q_m не оказываются в точности обратно пропорциональными друг другу.

Таким образом, в реальной ситуации можно рассчитывать лишь на выполнение условия:

$$\left| p_k - \frac{1}{q_m} \right| \leq \varepsilon, \quad (10)$$

означающего что для некоторого собственного числа p_k матрицы Z_E нашлось «почти обратное» к нему собственное число q_m матрицы Z_I , и, значит, p_k , скорее всего, является истинным полюсом сигнала.

Параметр ε фактически указывает, насколько в данном сигнале (при конкретном уровне присутствующего в нем шума) можно отличаться двум этим оценкам, он должен зависеть от уровня шума. Наши численные эксперименты показывают, что обычно значения ε находятся в диапазоне $0,01 \div 0,1$.

3. Численный эксперимент

Во многих задачах (например, в радиолокации) важной проблемой является обнаружение сигнала в шуме (как его времени прихода, так и его параметров). В связи с этим интересно рассмотреть работу предложенного метода для модельного сигнала следующего вида. Сигнал длительностью 0,1 с. (рис. 2) представляет собой последовательность затухающих синусоид, возникающих в случайные моменты времени.

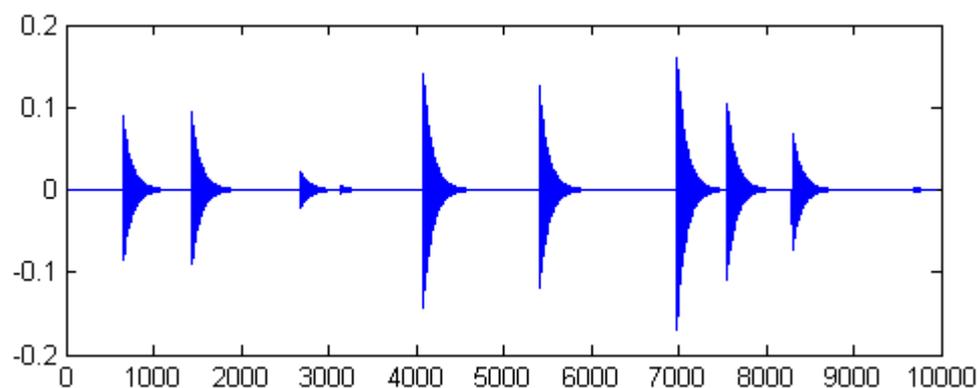


Рис. 2. Случайная последовательность синусоид

На рассматриваемом отрезке сигнала затухающая синусоида частотой 30 кГц возникла 10 раз (некоторые синусоиды частично перекрылись друг с другом). Амплитуда и

фаза синусоид — различны, частота и коэффициент затухания всех синусоид — одинаковы. Частота дискретизации сигнала — 100 кГц.

Зашумленный сигнал (в качестве шума взяты отсчеты нормально распределенной случайной величины) представлен на рис. 3. Отношение сигнал/шум (SNR) для данного сигнала будем вычислять по формуле:

$$SNR = 10 \log \frac{\frac{1}{M} \sum_{s=1}^M \frac{1}{N} \sum_{k=\eta_s}^{\eta_s+N} y_k^2}{\sigma^2}, \quad (11)$$

Здесь M — число импульсов на рассматриваемом временном отрезке сигнала, N — число отсчетов, за которое синусоида затухает до уровня 5 % от энергии в момент возникновения импульса, η_s — номер отсчета, соответствующий времени возникновения затухающей синусоиды, y_k — значения выходного сигнала в отсутствие шума, а σ — дисперсия шумового сигнала.

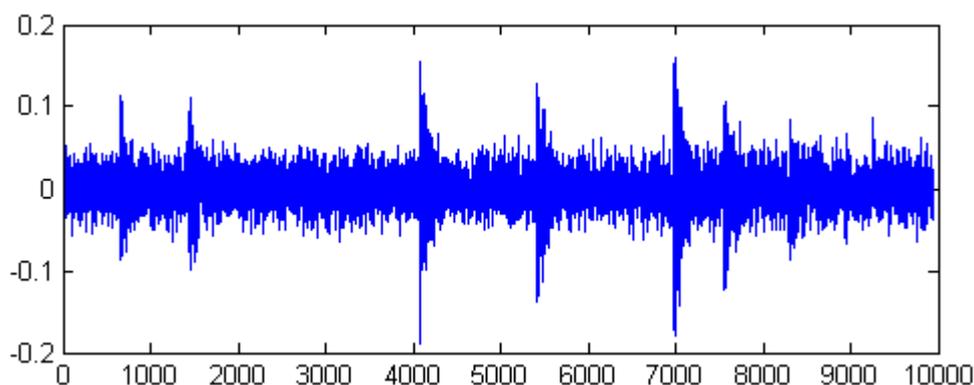


Рис. 3. Зашумленный сигнал

Для сигнала на рис. 3 значение SNR составило 7 дБ. Со сдвигом на 1 отсчет, предложенным в статье методом были обработаны сегменты сигнала длиной 0,5 мс каждый. Значение параметра ε было выбрано равным 0,01. Найденные значения частот показаны на рис. 4. Видно, что на участках с полезным сигналом модифицированным методом матричных пучков была обнаружена частота около 30 кГц. Важно, что на участках сигнала без полезной составляющей никаких частот не было найдено. Таким образом, предложенный метод позволяет найти время прихода сигнала и даже оценить его частоту.

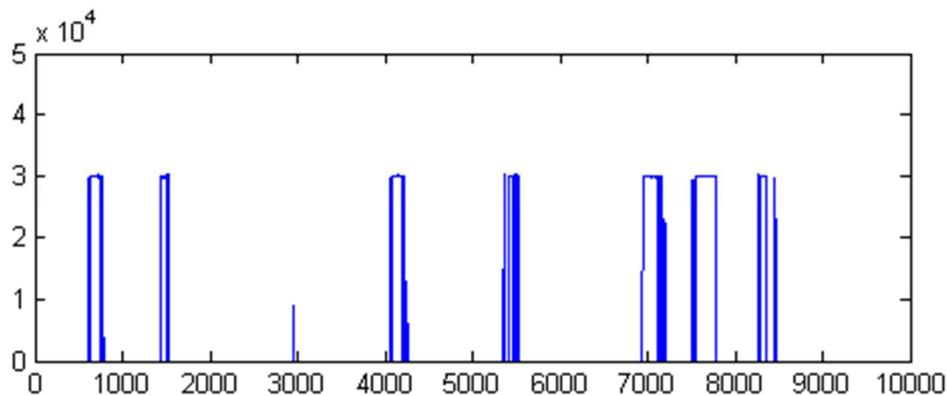


Рис. 4. Зависимость найденной частоты от отсчетов сигнала.

Заметим, что классический метод матричных пучков будет подстраивать шум под сумму синусоид и находить различные частоты на протяжении всего сигнала.

Сравним классический и модифицированный ММП. Для этого рассмотрим отрезок сигнала с 4100 по 4150 отсчет, в котором присутствует затухающая синусоида с частотой 30 кГц и «чисто шумовой» отрезок сигнала с 6000 по 6050 отсчет. Результаты сравнения сигналов представлены в табл. 1 и 2.

Таблица 1

Результаты обработки сигнала с полезной составляющей с частотой 30 кГц

	Классический метод матричных пучков	Модифицированный метод матричных пучков
Полюса	$-0,3138 + 0,9596i$ $0,8366 + 0,4215i$	$-0,3130 + 0,9354i$
Частоты (кГц)	30,223 7,428	30,140

Таблица 2

Результаты обработки сигнала без полезной составляющей (шума)

	Классический метод матричных пучков	Модифицированный метод матричных пучков
Полюса	$0,0518 + 0,7307i$ $0,7149 + 0,5868i$ $0,1290 + 0,0000i$	—
Частота (кГц)	23,874 10,938 0	—

Как можно видеть, для первого отрезка сигнала классический метод, помимо 30 кГц, нашел лишнюю частоту. Для второго отрезка классический метод также нашел ряд частот, подстроив под данный шум сумму комплексных экспонент.

Таким образом, ясно, что классический ММП не позволит определить время прихода полезного сигнала.

Заключение

В статье рассмотрен параметрический метод спектрального анализа — метод матричных пучков. Представлена его модификация, использующая помимо полюсов сигнала обратные к ним. При использовании предложенного метода вывод делается на основе поиска соответствующих пар полюсов. Проведено численное моделирование. Результаты показывают, что предложенный метод, в отличие от классического, способен выделять полюса, соответствующие полезному сигналу и определять время прихода полезного сигнала и его частоту.

Хотя в настоящее время имеется достаточно много методов, позволяющих решить задачу оценки параметров экспоненциально затухающего синусоидального сигнала (метод Прони, метод максимального правдоподобия, различные методы фильтрации), минимальные требования к объему априорной информации и возможность непосредственного определения всех параметров сигнала, позволяет говорить о перспективности предложенного метода для использования в алгоритмах идентификации систем и оценке их состояния, в частности для обработки сигналов с кориолисового расходомера при его работе в многофазных средах.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации, проект «Разработка отечественного массового кориолисового расходомера для нефтегазовой промышленности с функцией измерения расхода многофазных потоков» (соглашение № 14.578.21.0191 от 3.10.2016, идентификатор исследования RFMEFI57816X0191).

Авторы выражают благодарность профессору А.Л. Шестакову за постановку задачи и научное руководство.

Литература

1. Hua Y., Sarkar T.K. Matrix Pencil Method for Estimating Parameters of Exponentially Damped Undamped Sinusoids in Noise // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1990. Vol. 38, No. 5. P. 814–824. DOI: 10.1109/29.56027.
2. Potts D., Tasche M. Parameter Estimation for Non Increasing Exponential Sums by Prony-like Methods // Special Section on Statistical Signal & Array Processing. 2010. Vol. 90, P. 1631–1642. DOI: 10.1016/j.sigpro.2009.11.012.
3. Lin Y., Hodkinson P., Ernst M., Pines A. A Novel Detection-estimation Scheme for Noisy NMR Signals: Applications to Delayed Acquisition Data // Journal of Magnetic Resonance. 1997. Vol.128, P. 30–41. DOI: 10.1006/jmre.1997.1215.
4. Коновалюк М.А., Кузнецов Ю.В., Баев А.Б. Идентификация объектов сложной формы в сверхкороткоимпульсной радиолокации // III Всероссийская конференция «Радиолокация и связь» (Москва, 26–30 октября 2009 г.) Москва: Изд-во ИПЭ им. В.А.Котельникова РАН, 2009. С. 932–936.
5. Персичкин А.А., Шпилевой А.А. О методике оценки параметров сейсмических сигналов // Вестник Балтийского федерального университета им. И. Канта. 2015. № 10. С. 122–125.
6. Bhuiyan M., Malyarenko E.V., Pantea M.A., Capaldi D., Baylor A.E., Maev R.Gr. Time-frequency Analysis of Clinical Percussion Signals Using Matrix Pencil Method // Journal of Electrical and Computer Engineering 2015. Vol. 2015. P. 340–347. DOI: 10.1155/2015/274541.
7. Марпл-мл С.Л. Цифровой спектральный анализ и его приложения: пер. с англ. Москва: Изд-во Мир, 1990, 584 с.
8. Шестаков А.Л., Семенов А.С., Ибряева О.Л. Оценка несущей частоты случайной последовательности импульсов методом Прони // Вестник Южно-Уральского государственного университета. Серия: Математическое моделирование и программирование. 2009. № 37(170). С. 106–115.

9. Шостак С.В., Бакланов Е.Н., Стародубцев П.А., Шевченко А.П. Решение задачи «обнаружение-измерение дальности» для малоподвижных объектов методом активной корреляции // Журнал Радиоэлектроники 2015. № 3. С. 101–117.
10. Логинов А.А., Морозов О.А., Сорохтин Е.М., Сорохтин М.М., Реализация алгоритма поиска сигнала заданной формы на фоне шумов // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Физика твердого тела. 2005. № 1(18). С. 141–145.
11. Ван Трис Г. Теория обнаружения, оценок и модуляции. Москва: Изд-во Сов. радио, 1977. 650 с.
12. Stoica P., Moses R.L., Friedlander B., Soderstrom T. Maximum Likelihood Estimation of the Parameters of Multiple Sinusoids from Noisy Measurements // IEEE Transaction on Acoustics, Speech and Signal Processing. 1989. Vol. 37, No. 3. P. 378–392. DOI: 10.1109/29.21705.
13. Yang X., Huang B., Gao H. A Direct Maximum Likelihood Optimization Approach to Identification of LPV Time-delay Systems // Journal of the Franklin Institute. 2016. Vol. 353. P. 1862–1881. DOI: 10.1016/j.jfranklin.2016.03.005.
14. Hua Y., Sarkar T.K. On the Total Least Squares Linear Prediction Method for Frequency Estimation // IEEE Transaction on Acoustics, Speech and Signal Processing. 1990. P. 2186–2189. DOI: 10.1109/29.61547.
15. Гантмахер Ф. Р. Теория матриц. — 5-е изд. Москва: ФИЗМАТЛИТ, 2004. 560 с.
16. Penrose R. On Best Approximate Solutions of Linear Matrix Equations // Mathematical Proceedings of the Cambridge Philosophical Society. 1956. Vol. 52. P. 17–19. DOI: 10.1017/s0305004100030929.
17. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. Москва: Высш. шк, 2000. 480 с.

Ибряева Ольга Леонидовна, к.ф.-м.н., доцент, кафедра прикладной математики и программирования, Институт естественных и точных наук, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Салов Данил Дмитриевич, студент, Высшая школа электроники и компьютерных наук, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

DOI: 10.14529/cmse170102

MODIFICATION OF THE MATRIX PENCIL METHOD USING A COMBINED EVALUATION OF SIGNAL POLES AND THEIR INVERSES

© 2017 O.L. Ibryaeva, D.D. Salov

South Ural State University (pr. Lenina 76, Chelyabinsk, 454080 Russia),

E-mail: ibryaevaol@susu.ru, salovdd@yandex.ru

Received: 20.10.2016

The Prony-like parametric method of signal processing, namely the Matrix Pencil Method, is considered in the paper. The method is able to find frequencies, damping factors, phases and amplitudes of a sinusoids sum. It needs fewer number of operations than Prony method and hence lower computational error. A modification of the method using a combined evaluation of signal poles and their inverses is proposed. This modification is able to solve the problem of true/false poles separation. Two matrix pencils with eigenvalues coinciding (in the absence of noise) with signal poles and their inverses are constructed from the signal samples. In case of noisy signal true/false poles separation is performed by: 1) SVD; 2) excessive order of prediction; 3) analysis of eigenvalues of two matrix pencils. Algorithms of the modified and classical Matrix Pencil Methods are given and compared on the example of signal detection in noise. It is shown that the classical method is not able to detect the time of arrival of the signal since it fits an exponential sum to the noise. The modified method can detect both the time of arrival and the signal frequency. The proposed algorithm of signal detection is suitable for use with signals of sufficiently general form (sum of decaying sine waves) and does not require distribution laws of signal and its noise component, as the maximum likelihood method.

Keywords: signal processing, matrix pencil method, detection of signal in noise, unknown time of arrival, frequency estimation.

FOR CITATION

Ibryaeva O.L., Salov D.D. Modification of the Matrix Pencil Method using a combined evaluation of signal poles and their inverses. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2017. vol. 6, no. 1. pp. 26–37. (in Russian) DOI: 10.14529/cmse170102.

References

1. Hua Y., Sarkar T.K. Matrix Pencil Method for Estimating Parameters of Exponentially Damped Undamped Sinusoids in Noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1990. vol. 38, no. 5. pp. 814–824. DOI: 10.1109/29.56027.
2. Potts D., Tasche M. Parameter Estimation for Non Increasing Exponential Sums by Prony-like Methods. *Special Section on Statistical Signal & Array Processing*. 2010. vol. 90, pp. 1631–1642. DOI: 10.1016/j.sigpro.2009.11.012.
3. Lin Y., Hodkinson P., Ernst M., Pines A. A Novel Detection-estimation Scheme for Noisy NMR Signals: Applications to Delayed Acquisition Data. *Journal of Magnetic Resonance*. 1997. vol. 128, pp. 30–41. DOI: 10.1006/jmre.1997.1215.
4. Konovalyuk M.A., Kuznetsov Yu.V., Baev A.B. Identification of Objects Having Complex form in a Super Short Impulsive Radiolocation. *III Vserossiyskaya konferentsiya "Radiolokatsiya i svyaz" (Moskva, 26–30 oktyabrya 2009)* [Radiolocation and Radio Service: Reports of 3rd All-Russia Technological Conference "Radiolocation and Radio Service" (Moscow, Russia, October, 26–30, 2009)]. Moscow, Publishing of IRE V.A. Kotel'nikova RAN, 2010. pp. 932–936. (in Russian)
5. Persichkin A.A., Shpilevoj A.A. About the Methodology for Assessing the Parameters Seismic Signals. *Vestnik Baltiyskogo federal'nogo universiteta im. I. Kanta* [Vestnik Immanuel Kant Baltic Federal University]. 2015. no 10. pp. 122–125. (in Russian)
6. Bhuiyan M., Malyarenko E.V., Pantea M.A., Capaldi D., Baylor A.E., Maev R.Gr. Time-frequency Analysis of Clinical Percussion Signals Using Matrix Pencil Method. *Journal of Electrical and Computer Engineering* 2015. vol. 2015. pp. 340–347. DOI: 10.1155/2015/274541.
7. Marple, Jr.S.L. *Digital Spectral Analysis with Applications*. New Jersey, Prentice-Hall, Inc, Englewood Cliffs, 1987. p 584.

8. Shestakov A.L., Semenov A.S., Ibryaeva O.L. Carrier Frequency Estimation for Random Pulse Train Using Prony's Method. *Vestnik Yuzhno-Uralskogo gosudarstvennogo universiteta. Seriya: Matematicheskoe modelirovanie i programmirovaniye* [Bulletin of South Ural State University. Series: Mathematical Modelling, Programming & Computer Software]. 2009. vol. 37(170). pp. 106-115. (in Russian)
9. Shostak S.V., Baklanov E.N., Starodubtsev P.A., Shevchenko A.P. Solution of the Problem "Detection Range Measurement" for Low-mobility Objects by Methods of Active Correlation. *Zhurnal Radioelektroniki* [Journal of Radio Electronics]. 2015. vol. 3. pp. 101-117. (in Russian)
10. Loginov A.A., Morozov O.A., Sorokhtin E.M., Sorokhtin M.M. Synthesis Algorithm Identify a Given Waveform on Background Noise. *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo. Seriya: Fizika tverdogo tela*. [Vestnik of Lobachevsky University of Nizhni Novgorod. Series: Solid State Physics]. 2005. Vol. 1(18). pp. 141-145. (in Russian)
11. Van Tris G. *Teoriya obnaruzheniya, otsenok i modulyatsii* [Detection, Estimation, and Modulation]. Moscow, Publisher Soviet radio, 1977. 650 p.
12. Stoica P., Moses R.L., Friedlander B., Soderstrom T. Maximum Likelihood Estimation of the Parameters of Multiple Sinusoids from Noisy Measurements. *IEEE Transaction on Acoust. Speech, Signal Process.* 1989. vol. 37, no. 3. pp. 378-392. DOI: 10.1109/29.21705.
13. Yang X., Huang B., Gao H. A Direct Maximum Likelihood Optimization Approach to Identification of LPV Time-delay Systems. *Journal of the Franklin Institute.* 2016. vol. 353. pp. 1862-1881. DOI: 10.1016/j.jfranklin.2016.03.005.
14. Hua Y., Sarkar T.K. On the Total Least Squares Linear Prediction Method for Frequency Estimation. *IEEE Transaction on Acoustics, Speech and Signal Processing.* 1990. pp. 2186-2189. DOI: 10.1109/29.61547.
15. Gantmakher F. R. *Teoriya matrits* [The Theory of Matrices]. Moscow: Fizmatlit. 2004. 560 p.
16. Penrose R. On Best Approximate Solutions of Linear Matrix Equations. *Mathematical Proceedings of the Cambridge Philosophical Society.* 1956. vol. 52. pp. 17-19. DOI: 10.1017/s0305004100030929.
17. Ventcel E.S., Ovcharov L.A. *Teoriya veroyatnostey i yeye inzhenernyye prilozheniya* [Theory of Probability and its Engineering Applications], Moscow: Higher School, 2000. 480 p.

ЭКСПЕРИМЕНТАЛЬНОЕ СРАВНЕНИЕ АЛГОРИТМОВ В ПАРАЛЛЕЛЬНОМ МЕТОДЕ ВЛОЖЕННЫХ СЕЧЕНИЙ*

© 2017 г. А.Ю. Пирова, Н.Ю. Кудрявцев, И.Б. Мееров

Нижегородский государственный университет им. Н.И. Лобачевского

(603022, г. Нижний Новгород, просп. Гагарина, д. 23)

E-mail: anna.pirova@itmm.unn.ru, n.yu.kudriavtsev@gmail.com,

iosif.meyerov@itmm.unn.ru

Поступила в редакцию: 20.10.2016

В прямых методах решения больших разреженных систем линейных алгебраических уравнений применяется процедура переупорядочения строк и столбцов исходной матрицы. Целью данной процедуры является сокращение числа ненулевых элементов в процессе последующей численной факторизации. Нахождение перестановки, минимизирующей число ненулевых элементов в факторе, является NP-полной задачей. Для решения этой задачи применяются эвристические методы. Результаты применения данных методов могут быть оценены как с точки зрения качества получаемых перестановок (заполнение фактора матрицы после переупорядочения), так и с точки зрения временных затрат на построение перестановок. Многоуровневый метод вложенных сечений, показывающий достаточно хорошие результаты по обоим критериям, является одним из наиболее распространенных методов переупорядочения. Метод имеет определенные ресурсы внутреннего параллелизма, активно используемые в ряде реализаций (ParMETIS, mtMETIS, PT-SCOTCH, PMORSy). Вместе с тем, низкая арифметическая интенсивность, нерегулярный доступ к памяти, дисбаланс вычислительной нагрузки и необходимость поиска компромисса между временем работы и качеством перестановок мотивируют дальнейшие исследования метода.

В данной работе выполняется сравнение ряда алгоритмов, применяемых на разных этапах метода вложенных сечений, с точки зрения их влияния на заполнение фактора и время работы в параллельном случае. Реализация алгоритмов и эксперименты выполнены в рамках ранее разработанной параллельной библиотеки переупорядочения матриц PMORSy, опережающей аналоги на ряде матриц коллекции университета Флориды. В результате выполненной работы удалось выделить наиболее перспективную комбинацию алгоритмов и улучшить качество перестановок и время работы PMORSy.

Ключевые слова: метод вложенных сечений, переупорядочение разреженных матриц, параллельный алгоритм.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Пирова А.Ю., Кудрявцев Н.Ю., Мееров И.Б. Экспериментальное сравнение алгоритмов в параллельном многоуровневом методе вложенных сечений // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. Т. 6, № 1. С. 38–55. DOI: 10.14529/cmse170103.

Введение

Решение систем линейных уравнений прямыми методами заключается в разложении (факторизации) матрицы системы на произведение верхне- и нижнетреугольной матриц и дальнейшего решения двух систем линейных уравнений с полученными треугольными матрицами. При этом, в случае разреженной матрицы системы, возникает проблема *заполнения*: число ненулевых элементов фактора матрицы в разы больше числа ненуле-

* Статья рекомендована к публикации программным комитетом Международной научной конференции «Параллельные вычислительные технологии (ПаВТ) 2017».

вых элементов исходной матрицы. Чтобы снизить заполнение фактора, предварительно применяется процедура *переупорядочения* строк и столбцов исходной матрицы. От результатов переупорядочения зависят затраты памяти на хранение фактора и времени на выполнение дальнейшей факторизации. Поэтому под *качеством перестановки* чаще всего понимают число ненулевых элементов факторизованной матрицы (чем меньше, тем лучше). Кроме того, перестановка влияет на потенциал распараллеливания наиболее трудоемкого этапа решения — численной факторизации, поскольку она определяет взаимное расположение ненулевых элементов.

Нахождение перестановки, минимизирующей число ненулевых элементов в факторе, является NP-полной задачей [18]. На практике для ее решения применяются две группы эвристических методов — методы минимальной степени [17] и методы вложенных сечений [5, 6]. Известно, что перестановки, полученные методом вложенных сечений, имеют больший потенциал для параллелизма на этапе численной факторизации [4]. Поэтому в данной работе для реализации была выбрана модификация этого метода — многоуровневый метод вложенных сечений [1, 8], применяющийся с середины 1990-х годов.

В настоящее время в мире разработан ряд библиотек для переупорядочения разреженных матриц, преимущественно имеющих последовательные реализации или реализации для систем с распределенной памятью. Многоуровневый метод вложенных сечений, в частности, реализован в свободно распространяемых библиотеках METIS [10] и Scotch [12], широко используемых в решателях СЛАУ, а также в их аналогах для систем с распределенной памятью ParMETIS [7] и PT-Scotch [2]. Среди российских разработок отметим пакет МАТРУЗ [20]. В последние годы ведется работа по разработке реализаций многоуровневого метода вложенных сечений для систем с общей памятью. Так, в 2013 году авторами Scotch обсуждалось распараллеливание отдельных этапов многоуровневого метода [13], в 2015 году авторами METIS был представлен алгоритм переупорядочения для систем с общей памятью, реализованный в пакете mt-metis [11]. В то же время в Нижегородском государственном университете им. Н.И. Лобачевского была разработана библиотека PMORSy [15, 19], также ориентированная на многоядерные системы с общей памятью. На матрицах из разных предметных областей было показано [15], что реализация из PMORSy сопоставима по времени и качеству с реализациями из ParMETIS, mt-metis и PT-Scotch. При этом на ряде матриц PMORSy опережает аналоги, на других — проигрывает им в смысле указанных выше критериев.

Отдельного обсуждения заслуживает вопрос об эффективности распараллеливания. Известно, что алгоритмы переупорядочения обладают рядом особенностей, существенно затрудняющих возможность создания хорошо масштабируемого параллельного алгоритма. Среди таких особенностей можно отметить достаточно низкую арифметическую интенсивность, нерегулярный доступ к памяти, дисбаланс вычислительной нагрузки при решении независимых подзадач. Указанные причины приводят к недостаточной эффективности масштабируемости параллельных алгоритмов во всех указанных выше библиотеках, что обуславливает актуальность продолжения исследований в данной области.

Основная цель данной работы заключалась в изучении множества отдельных алгоритмов, предложенных в разное время для решения задач, возникающих на ряде этапов метода вложенных сечений. Исследовалось влияние данных алгоритмов на качество получаемых перестановок и время работы не только в последовательном [8, 9], но и, глав-

ное, в параллельном случае. В качестве базовой реализации использовалась библиотека PMORSy, тестовые задачи брались из коллекции университета Флориды.

Статья построена следующим образом. В разделе 1 приводится описание метода вложенных сечений и параллельного алгоритма, использующегося в PMORSy. В разделе 2 описаны отдельные этапы многоуровневого метода и особенности их программной реализации. В разделе 3 приводятся результаты сравнения времени и качества PMORSy в зависимости от выбора рассматриваемых алгоритмов, а также приводится сравнение с библиотекой mt-metis. В заключении обобщены результаты работы и приведены планы дальнейших исследований.

1. Параллельный многоуровневый метод вложенных сечений

Приведем краткое описание многоуровневого метода вложенных сечений. Базовым для метода является понятие *разделителя* — множества вершин, удаление которого расщепляет граф на две или более несвязные части. Процесс нахождения перестановки выполняется следующим образом. В графе, построенном по матрице СЛАУ, находят разделитель, который делит граф на две близкие по размеру части. Затем разделитель удаляется из графа, его вершины нумеруются последними свободными индексами, а процесс продолжается независимо на образовавшихся несвязных подграфах. Переупорядочение завершается, когда все вершины графа занумерованы. Качество переупорядочения зависит от того, насколько удачно были вычислен разделитель на каждом шаге. Чем меньше разделитель и дисбаланс образовавшихся после его удаления частей, тем лучше будет найденная перестановка.

При использовании многоуровневого метода вложенных сечений разделитель вычисляется в три этапа: огрубление, разделение, развертывание. На этапе огрубления по исходному графу с помощью паросочетаний строится последовательность графов меньшего размера, в которой каждый следующий граф представляет собой огрубленную структуру предыдущего графа. На этапе разделения выполняется нахождение разделителя и несвязных частей самого маленького в последовательности графа. Затем, в процессе развертывания, этот разделитель проецируется через всю последовательность графов на исходный граф. При этом используются алгоритмы улучшения разделения, которые уменьшают разделитель и сокращают дисбаланс образовавшихся несвязных частей. Псевдокод нахождения разделителя многоуровневым методом приведен на рис. 1. Подробное описание последовательного алгоритма переупорядочения, реализованного в PMORSy, приведено в [14].

Распараллеливание многоуровневого метода вложенных сечений в PMORSy использует принцип «разделяй и властвуй», заложенный в методе вложенных сечений. Алгоритм переупорядочения представляет собой параллельную обработку очереди независимых задач. Задачей, назначаемой потоку, является вычисление разделителя некоторого графа и получение новых несвязных подграфов. После того, как разделитель графа найден, поток создает новые задачи по обработке порожденных подграфов, если они достаточно большие, и выполняет рекурсивную обработку маленьких подграфов. Для реализации алгоритма использовался механизм логических задач OpenMP 3.0.

```

1.  (S0, G1*, ..., Gk*) NDStep(graph G0(V0, E0), int nSt,
2.      int sm, int r) {
3.      if |V0| < sm {
4.          S0 = AutomaticNestedDissection(G0);
5.          return (S0, NULL);
6.      }
7.      else {
8.          i = 0;
9.          while (i < nSt) && (Vi > sm) {
10.             Gi+1(Vi+1, Ei+1) = Coarse(Gi(Vi, Ei)); i++;
11.          }
12.          m = i;
13.          Pm(Sm, Vm,1, Vm,2) = InitializePartition(Gm, r);
14.          for(i = m; i >= 0; i--) {
15.              Pi-1 = ProjectPartition(Pi, Gi);
16.              RefinePartition(Pi-1);
17.          }
18.          (G1*, G2*, ..., Gk*) = FindSubgraphs(G0, P0);
19.          return (S0, G1*, G2*, ..., Gk*);
20. }

```

Рис. 1. Многоуровневый метод вложенных сечений в PMORSy [14]

Псевдокод параллельного алгоритма переупорядочения приведен на рис. 2. Основная функция *PMORSyOrdering* принимает на вход портрет графа матрицы G и параметры последовательного и параллельного алгоритмов nSt , sm , r , b . Задаче, назначаемой одному потоку, соответствует вызов функции *NDStepParallel*, которая принимает на вход подграф G_0 и параметры алгоритмов, и возвращает найденный разделитель S . После того, как для всех подграфов был найден разделитель, выполняется их слияние в массив перестановки.

```

1.  int* NDStepParallel(graph G0(V0, E0), int nSt, int sm,
2.      int r, int b) {
3.      (S0, G1*, G2*, ..., Gk*) = NDStep(G0, nSt, sm, r);
4.      for (i = 0; i < k; i++) {
5.          #pragma omp task if (Gi* > b)
6.          NDStepParallel(Gi*, nSt, sm, r, b);
7.      }
8.      return S;
9.  }
10.
11. void PMORSyOrdering(graph G, int* iperm,
12.     int nSt, int sm, int r, int b) {
13.     #pragma omp parallel
14.     #pragma omp single
15.     S = NDStepParallel(G, nSt, sm, r, b);
16.     iperm = Merge(S);
17. }

```

Рис. 2. Параллельный метод вложенных сечений в PMORSy [14]

2. Алгоритмы многоуровневого метода вложенных сечений

Рассмотрим подробнее алгоритм получения разделителя многоуровневым методом (рис. 1). Функция *NDStep* на вход получает граф G_0 и параметры алгоритма: nSt (мак-

симальное число шагов огрубления), sm (минимальный размер графа, к которому применяется многоуровневый метод вложенных сечений), r (число перезапусков алгоритма разделения). Выход функции *NDStep* — разделитель S_0 и новые подграфы G_1^*, \dots, G_k^* . Этапу огрубления соответствует вызов в цикле функции *Coarse* (строки 8–10), этапу разделения — вызов функции *InitializePartition* (строка 13), этапу развертывания — вызов в цикле функций *ProjectPartition* и *RefinePartition* (строки 14–17).

Для выполнения каждого из этапов многоуровневого метода разработан ряд последовательных алгоритмов, успешно применяющихся в рамках последовательного и параллельного метода вложенных сечений.

Так, для огрубления используются алгоритмы поиска максимального паросочетания. Для этого применяются алгоритмы случайных паросочетаний (random matching, RM), паросочетания тяжелых ребер (heavy edge matching, HEM), паросочетания легких ребер (light edge matching, LEM), паросочетания тяжелых клик (heavy clique matching, HCM) и др. Их описание можно найти, например, в работе [8]. Предварительные эксперименты показали, что паросочетания легких ребер или, аналогично, легких вершин, приводят к образованию слишком плотных графов, что снижает итоговое качество переупорядочения. Поэтому в данной работе были выбраны алгоритм случайных паросочетаний (RM), паросочетания тяжелых ребер (HEM), паросочетание тяжелых клик (HCM).

На этапе разделения выполняется определение вершинного разделителя графа и несвязных частей, образующихся после его удаления. Часто вершинный разделитель находят из реберного разделителя. Для этого изначально разделяют все вершины на два множества, близких по числу вершин или по их суммарному весу. Затем применяют алгоритм поиска минимальной вершинной оболочки либо включают в вершинный разделитель все вершины, инцидентные реберному разделителю.

Одним из простых методов получения вершинного разделителя является выделение среднего уровня смежности поиском в ширину, запущенным из псевдопериферийной вершины (level structure from pseudoperipheral vertex, LS) [6]. Для получения реберного разделителя также используются методы, основанные на поиске в ширину из некоторой вершины, например, алгоритм растущего разделения (graph growing partition, GG), алгоритм жадного растущего разделения (greedy graph growing partition, GGG). Алгоритм Кернигана—Лина (Kernighan–Lin partition, KL) выполняет минимизацию реберного разделителя для некоторого начального разделения, поэтому может использоваться как процедура, улучшающая свойства заданного разделения. Также для разделения графа применяются методы, использующие спектральную или геометрическую информацию о графе. Однако предварительные эксперименты показали, что их использование в рамках переупорядочения ведет к большим временным затратам при том, что качество полученных с их помощью перестановок не отличается значительно от качества перестановок, полученных другими методами. Поэтому в данной работе эти методы не рассматривались. Обзор алгоритмов реберного разделения графа можно найти в [8, 16].

Все алгоритмы, использованные в данной работе, были реализованы по описанию из [8]. Для алгоритмов, получающих реберный разделитель, в вершинный разделитель включались все вершины, инцидентные реберному разделителю. Кроме того, эти алгоритмы запускались несколько раз из случайных начальных вершин, чтобы улучшить качество получаемого разделения. Алгоритм Кернигана—Лина в качестве начального

разделения получал результаты растущего разделения графа. Ко всем алгоритмам на этапе развертывания графа применялась процедура улучшения вершинного разделения, описанная в [14].

3. Результаты вычислительных экспериментов

3.1. Методика проведения экспериментов

Основная цель, поставленная в данной работе, — определить, какие алгоритмы огрубления и разделения, необходимые для многоуровневого метода вложенных сечений, а также параметры самой многоуровневой схемы, целесообразно использовать в рамках имеющейся параллельной реализации переупорядочения в PMORSy. Результаты переупорядочения традиционно оцениваются с точки зрения качества перестановок (то есть числа ненулевых элементов фактора) и времени работы. При этом подбор параметров для улучшения одного из этих критериев, нередко приводит к ухудшению второго. В данной серии экспериментов не ставилось задачи найти лучшие перестановки по времени и (или) качеству для каждой из рассмотренных матриц. Напротив, для всех тестовых задач использовались одни и те же «компромиссные» значения основных параметров метода.

Вычислительные эксперименты проводились на 48 матрицах из коллекции Университета Флориды [3], из разных предметных областей. Все матрицы были симметричные, вещественные, порядком от $2,2 \times 10^6$ до $2,8 \times 10^8$, с числом ненулевых элементов от $7,6 \times 10^6$ до $7,6 \times 10^9$. В тестовый набор вошли 28 самых больших симметричных матриц коллекции. Все эксперименты проводились на узле суперкомпьютера Лобачевский, имеющем следующие характеристики: два 8-ядерных процессора Intel Sandy Bridge E5-2660, частота 2.2 GHz, память 64 GB, операционная система Linux CentOS 6.4. Использовался компилятор Intel C++ Compiler и генератор случайных чисел из библиотеки Intel MKL, входящие в пакет Intel Parallel Studio XE 2015 Cluster Edition.

Для контроля результатов PMORSy проводилось сравнение с библиотекой *mt-metis v.0.5.0* [11]. Данная библиотека предназначена для систем с общей памятью, как и PMORSy, и входит в число лучших реализаций многоуровневого метода вложенных сечений. Более полное сравнение с библиотеками *ParMETIS* и *mt-metis* приведено в работе [15].

В рамках многоуровневого метода вложенных сечений в PMORSy сравнивались следующие алгоритмы огрубления: случайных паросочетаний (RM), паросочетания тяжелых ребер (HEM), паросочетание тяжелых клик (HCM). Для каждого из них были рассмотрены алгоритмы разделения: поиск в ширину из псевдопериферийной вершины (LS), алгоритм растущего разделения (GG), алгоритм жадного растущего разделения (GGG), алгоритм Кернигана—Лина (KL). Кроме того, в экспериментах варьировались параметры многоуровневого метода, связанные с числом запусков алгоритмов огрубления и разделения. Были использованы следующие значения:

- размер минимального графа (sm): 20, 100;
- максимальное число итераций алгоритма огрубления (nSt): 10, 20, 100;
- число запусков алгоритма разделения для алгоритмов GG, GGG, KL (r): 5, 10.

Всего для каждой матрицы было проведено 126 экспериментов для фиксированного числа потоков, запуски производились в 1 и 16 потоков. Библиотека *mt-metis* запускалась с параметрами по умолчанию. Отметим, что и PMORSy, и *mt-metis* в нескольких

запусках с одними и теми же параметрами переупорядочения дают близкие по времени, но разные по качеству результаты. Это связано с тем, что в алгоритмах используются псевдослучайные числа, генерируемые на разных потоках. Как правило, разница в качестве переупорядочения на одном графе составляет не более 2%, поэтому в данной серии экспериментов запуски проводились однократно. Для каждого запуска собирались следующие данные: время и качество работы переупорядочения, время обработки самого большого подграфа, отношение времени и качества переупорядочения к результатам, полученным mt-metis.

Результаты экспериментов далее представлены в следующем виде. Все эксперименты были сгруппированы по комбинациям параметров размера минимального подграфа и максимального числа шагов округления (sm и nSt). Результаты переупорядочения, в которых качество было хуже, чем у mt-metis, более чем на 60%, не участвовали в сравнении времени. Не анализировались и результаты экспериментов, занявших более 25 минут. Полученные значения времени и качества для каждой матрицы были отображены на цветовую шкалу, на которой минимуму среди всех 126 запусков соответствует зеленый цвет, максимуму — красный цвет, промежуточным значениям — желтый цвет. Исключенным значениям соответствует белый цвет. Во всех таблицах матрицы приведены в порядке увеличения числа ненулевых элементов. Таким образом, наиболее вычислительно трудоемкие задачи сгруппированы внизу таблицы.

3.2. Сравнение алгоритмов при получении первого разделителя

Рассмотрим работу многоуровневого метода сечений при получении первого разделителя. Этот этап занимает около 15% общего времени последовательной версии переупорядочения для большинства матриц. Из этого времени по 40–60% времени выполняются этапы округления и улучшения разделения, около 3% — этап разделения. В PMORSy вычисление первого разделителя выполняется последовательно, и в параллельной версии занимает 30–55% общего времени работы переупорядочения. Это является существенным ограничением для масштабирования приложения на большое число потоков, особенно для графов с числом вершин более 10^7 . Для сокращения времени вычислений требуется распараллеливание отдельных этапов многоуровневого метода.

3.2.1. Алгоритмы округления

Рассмотрим работу алгоритмов округления при получении первого разделителя. Сравним графы, полученные различными алгоритмами округления после 10 шагов. Для большинства матриц (45 из 48) был получен граф с числом вершин, на три порядка меньшим исходного, и плотностью порядка 10^{-2} . В большинстве экспериментов граф меньшего размера был получен алгоритмами RM или HSM, при этом графы, полученные RM, были в 1,5–11,7 раз плотнее, чем полученные другими алгоритмами. Если продолжать округление далее, то для получения графа из 100–150 вершин алгоритмы выполняют почти одинаковое число шагов (разница 1–2 шага).

Во всех экспериментах 4 матрицы из 48 были недостаточно сжаты большинством алгоритмов. Это матрицы boyd2, HTC_336_9129, kkt_power, c_big. Причина такой работы алгоритмов — наличие в исходной матрице плотной строки, т.е. вершины графа матрицы, связанной почти со всеми другими вершинами.

Сравним время работы алгоритмов округления. При выполнении 10 шагов сжатия на большинстве графов наиболее быстро работал алгоритм НЕМ, наиболее медленно —

RM на графах размера до 10^7 , HCM — на графах размера более 10^7 . При огрублении до 100 вершин наиболее быстро работал алгоритм НЕМ, наиболее медленно — алгоритм RM (в 1,1–3 раза больше по сравнению с НЕМ). На четырех самых больших матрицах дольше всего работал алгоритм HCM.

Если сравнивать два наиболее часто используемых алгоритма — RM и НЕМ, то НЕМ позволяет за меньшее время получить граф большего, чем у RM, размера, но меньшей плотности. Соотношение числа вершин полученного графа составляет 1,2–2 раза в пользу RM, соотношение плотностей — 1,5–5 раз в пользу НЕМ для большинства графов. По времени НЕМ работает в среднем на 40% быстрее при 10 шагах огрубления, в 2 раза быстрее при 20 шагах огрубления и огрублении до размера 100. При этом для исходных графов порядка 10^6 не достаточно 10 итераций алгоритма огрубления для получения графа размером не более 1000.

3.2.2. Алгоритмы разделения

Сравним работу алгоритмов разделения в тех экспериментах, где граф был сжат до 100 вершин. В данном разделе рассмотрим алгоритмы GG, GGG, KL, которые изначально находят реберный разделитель.

По размеру вершинного разделителя худшие результаты были получены алгоритмом GG при 5 перезапусках, лучшие — GGG и KL при 10 перезапусках для графов, полученных HCM и НЕМ (наиболее разреженных). При этом для трети матриц дисбаланс разделений, полученный различными алгоритмами, различается в пределах 10%. Для большинства матриц лучший дисбаланс разделения был получен алгоритмами GGG и KL. Увеличение числа перезапусков в 2 раза меньше всего влияет на качество алгоритма GGG, для которого улучшение дисбаланса составляет в среднем 5%, улучшение размера разделителя — 1%. Наиболее целесообразно увеличивать число перезапусков алгоритма GG, для которого улучшение дисбаланса и размера разделителя в среднем составляет 5%.

Сравним время работы алгоритмов разделения на первой задаче. Алгоритмы GGG, KL работают на один порядок дольше, чем GG (4–60 раз в зависимости от плотности матрицы), при этом KL работает в 1,5–2,7 раз медленнее, чем GGG. При этом порядок времени работы алгоритмов разделения на сжатых графах 10^{-3} – 10^{-4} с. Увеличение в два раза числа перезапусков алгоритма GG увеличивает время его работы в среднем на 70%, но не более 5,2 раз; алгоритмов GGG и KL — в среднем в 2 раза, но не более 3,5 раз.

Алгоритм улучшения разделения является алгоритмом локальной оптимизации: он ищет минимум функции, зависящей от характеристик разделения — веса разделителя и дисбаланса частей. Для большинства графов в результате данной процедуры начальные разделения, полученные различными алгоритмами, были сведены к одинаковым разделениям. Исключение в основном составляют графы с вершиной большой степени. Поэтому можно рассматривать время, необходимое алгоритму улучшения разделения, как характеристику качества начального разделения графа — чем оно меньше, тем лучше.

В большинстве экспериментов наиболее быстро улучшение разделения выполнялось после разделения алгоритмом KL, наиболее медленно — после разделения LS и GGG. В 2/3 запусков для графов, полученных одним методом огрубления, время улучшения разделения после различных алгоритмов разделения различалось не более чем в 2 раза, в среднем — на 30%. Значительные различия во времени работы алгоритма улучшения

разделения наблюдались в случаях, когда граф, полученный огрублением, содержал вершину большой степени или был близким к плотному. Если сравнивать время работы улучшения разделения после 5 и 10 перезапусков алгоритмов разделения, то из экспериментов видно, что для 90% запусков разница во времени работы улучшения разделения была не более 1,5 раз, при этом в среднем разница составляла 4–13%.

3.3. Сравнение алгоритмов в последовательной версии

Вычислительные эксперименты, проведенные для библиотеки PMORSy, запущенной в 1 поток, показали следующее:

1. Часть экспериментов привела к перестановкам с неприемлемым размером фактора (значительно большим, чем в других запусках). Основная причина такой работы переупорядочения — некорректная работа алгоритмов разделения и улучшения разделения на почти плотных графах и графах, содержащих вершину со степенью, значительно больше средней. Для таких графов не подходит использованный метод получения вершинного разделителя по реберному.
2. Перестановки с лучшим качеством были получены преимущественно при использовании алгоритмов огрубления НЕМ и НСМ и любого метода разделения, а также огрублением RM с разделением LS. Алгоритмы НЕМ и НСМ меньше других чувствительны к плотности начального графа и наличию вершин с большой степенью.
3. Для графов со степенями вершин менее 10 и небольшим разбросом степеней приемлема любая комбинация алгоритмов огрубления и разделения. Для большинства из них полученные перестановки различались по качеству в пределах 10%.
4. На качество получаемых перестановок в большей степени влияет алгоритм улучшения разделения, чем алгоритмы начального разделения. Это объясняется тем, что этот алгоритм выполняет дискретную оптимизацию, которая для разных начальных разделений находит близкие локально-оптимальные «улучшенные» разделения. При этом время работы улучшения разделения для различных начальных разделений примерно одинаково.
5. Графы с вершиной, степень которой значительно больше средней, а также графы большого порядка чувствительны к большому числу итераций огрубления и низкому ограничению по размеру графа, поскольку это приводит к слишком плотным графам в процессе огрубления. Для таких графов лучшие результаты были получены при минимальном графе в 100 вершин.
6. Использование большего числа перезапусков алгоритма разделения для большинства матриц незначительно улучшает конечное качество перестановок (около 5%). Среди рассмотренных алгоритмов более чувствительны к увеличению числа перезапусков алгоритмы GG и KL. При этом время работы переупорядочения увеличивается в среднем на 12%. Поэтому большое число запусков алгоритма разделения предлагается использовать для графов, содержащих вершины со степенью, значительно больше средней.
7. По соотношению общего времени работы переупорядочения быстрее всего работали запуски с алгоритмом LS, затем GG (+8%), GGG (+12%), KL (+20%).
8. Среди комбинаций параметров минимального размера графа и максимального числа шагов огрубления наиболее стабильными с точки зрения качества и небольшого времени работы были параметры $nSt = 10$ и $sm = 100$ (табл. 1).

Таблица 1

Время и качество переупорядочения PMORSy при работе в 1 поток с параметрами $nSt = 10$, $sm = 100$. Результаты приведены по цветовой шкале, где минимальному значению соответствует зеленый цвет, максимальному — красный цвет, промежуточным значениям — желтый цвет. Результаты, дающие неприемлемое заполнение фактора, отмечены белым

Качество PMORSy										Матрица	Время PMORSy																													
RM LS	HEM LS	HCM LS	RM GG 10	RM GGG 10	RM KL 10	HEM GG 10	HEM GGG 10	HCM GG 10	HCM GGG 10		HEM KL 10	HEM GG 5	HCM GG 5	HEM KL 5	HCM GG 5	HCM GGG 5	RM LS	HEM LS	HCM LS	RM GG 10	RM GGG 10	RM KL 10	HEM GG 10	HEM GGG 10	HEM KL 10	HCM GG 10	HCM GGG 10	HEM KL 10	HEM GG 5	HCM GG 5	HEM KL 5	HCM GG 5	HCM GGG 5							
																HTC 336 9129																								
																Lin																								
																boyd2																								
																darcy003																								
																c-big																								
																helm2d03																								
																parabolic_fem																								
																CO																								
																offshore																								
																apache2																								
																ecology2																								
																ecology1																								
																tmt_sym																								
																Si87H76																								
																hood																								
																BenElechi1																								
																G3_circuit																								
																thermal2																								
																af_3_k101																								
																bmw3_2																								
																pwtk																								
																kkt_power																								
																nlpkkt80																								
																af_shell3																								
																af_shell9																								
																Ga41As41H72																								
																msdoor																								
																StocF-1465																								
																gsm_106857																								
																F1																								
																Fault_639																								
																inline_1																								
																Emilia_923																								
																boneS10																								
																ldoor																								
																bone010																								
																dielFilterV2real																								
																af_shell10																								
																Hook_1498																								
																Geo_1438																								
																Serena																								
																audikw_1																								
																dielFilterV3real																								
																nlpkkt120																								
																Flan_1565																								
																nlpkkt160																								
																nlpkkt200																								
																nlpkkt240																								

3.4. Сравнение алгоритмов в параллельной версии

При запусках PMORSy в 16 потоков сохраняются тенденции по соотношению времени и качества, полученного разными алгоритмами, которые наблюдались при запусках в 1 поток. Дополнительно отметим следующее:

1. Перестановки с лучшим качеством были получены при использовании округлений НЕМ и НСМ, как и в однопоточном случае. Для матриц с числом ненулевых элементов в пределах 25×10^6 лучшими по качеству были эксперименты с минимальным графом размера 20, для матриц с большим числом ненулевых элементов — с минимальным графом размера 100.
2. Время работы PMORSy зависит в большей степени от примененного алгоритма округления, чем от алгоритма разделения. Так, во всех группах экспериментов быстрее всего работали запуски с округлением НЕМ, затем — НСМ (+ 6%), дольше всех — RM (+ 25% в среднем). Это связано с тем, что алгоритм RM требует дополнительной генерации массива случайных чисел по числу ребер графа.
3. Время работы переупорядочения при изменении алгоритма разделения графа в среднем увеличивается на 10–20%. В большинстве экспериментов быстрее всего работало переупорядочение с алгоритмом LS или GG с 5 перезапусками. Например, при использовании алгоритма округления НЕМ на большинстве матриц переупорядочение быстрее всего работало с разделением GG с 5 перезапусками, затем — GGG и KL с 5 перезапусками (+ 4–6%) и LS (+ 5%), дольше других работали эксперименты с KL с 10 перезапусками (+ 10%).
4. Использование большего числа перезапусков алгоритма разделения для большинства матриц также не дает значительного улучшения качества (средний выигрыш составляет 6%). При этом время работы переупорядочения для большинства матриц увеличивается, в среднем на 7%.
5. PMORSy при запусках в 16 потоков показывает масштабируемость 2–3,5 раза на матрицах порядка до $5,0 \times 10^6$; 3,5–4,5 раза на матрицах большего размера. Наибольшие значения масштабируемости достигаются на запусках с разделением KL, которые, однако, для большинства матриц работают дольше, чем запуски с другими разделениями.
6. Среди комбинаций параметров минимального размера графа и максимального числа шагов округления наиболее стабильными с точки зрения качества и небольшого времени работы были параметры $nSt = 10$ и $sm = 100$, как и при запусках в 1 поток, а также комбинация $nSt = 100$ и $sm = 100$.

Сравнение качества и времени работы для комбинации $nSt = 10$ и $sm = 100$ приведено ниже (табл. 2).

Таблица 2

Время и качество переупорядочения PMORSy при работе в 16 потоков с параметрами $nSt = 10$, $sm = 100$. Результаты приведены по цветовой шкале, где минимальному значению соответствует зеленый цвет, максимальному — красный цвет, промежуточным значениям — желтый цвет. Результаты, дающие неприемлемое качество, отмечены белым

Качество PMORSy														Матрица	Время PMORSy																						
RM LS	HEM LS	HCM LS	RM GG 10	RM GGG 10	RM KL 10	HEM GG 10	HEM GGG 10	HEM KL 10	HCM GG 10	HCM GGG 10	HCM KL 10	HEM GG 5	HEM GGG 5		HEM KL 5	HCM GG 5	HCM GGG 5	RM LS	HEM LS	HCM LS	RM GG 10	RM GGG 10	RM KL 10	HEM GG 10	HEM GGG 10	HEM KL 10	HCM GG 10	HCM GGG 10	HCM KL 10	HEM GG 5	HEM GGG 5	HEM KL 5	HCM GG 5	HCM GGG 5			
																	HTC 336 9129																				
																	Lin																				
																boyd2																					
																darcy003																					
																c-big																					
																helm2d03																					
																parabolic_fem																					
																CO																					
																offshore																					
																apache2																					
																ecology2																					
																ecology1																					
																tmt_sym																					
																Si87H76																					
																hood																					
																BenElechi1																					
																G3_circuit																					
																thermal2																					
																af_3_k101																					
																bmw3_2																					
																pwtk																					
																kkt_power																					
																nlpkkt80																					
																af_shell3																					
																af_shell9																					
																Ga41As41H72																					
																msdoor																					
																StocF-1465																					
																gsm_106857																					
																F1																					
																Fault_639																					
																inline_1																					
																Emilia_923																					
																boneS10																					
																ldoor																					
																bone010																					
																dielFilterV2real																					
																af_shell10																					
																Hook_1498																					
																Geo_1438																					
																Serena																					
																audikw_1																					
																dielFilterV3real																					
																nlpkkt120																					
																Flan_1565																					
																nlpkkt160																					
																nlpkkt200																					
																nlpkkt240																					

3.5. Сравнение с библиотекой *mt-metis*

Сравним библиотеки PMORSy и *mt-metis*. По качеству переупорядочения для большинства матриц библиотеки показывают близкие результаты, отличающиеся в пределах 10% в пользу PMORSy или *mt-metis*, при этом соотношение качества близко для запусков с использованием разных алгоритмов. Исключение составляют матрицы, имеющие плотную строку (например, *kkt_power*, *c-big* и др.), на которых качество работы PMORSy не стабильно. Время работы PMORSy на 19 матрицах из 48 при этом меньше, чем у *mt-metis*, в 1,1–3 раза. На остальных матрицах в большинстве экспериментов PMORSy отстает от *mt-metis* в 1,1–3,5 раза. На самых больших матрицах из тестового набора PMORSy значительно отстает по времени от *mt-metis* (в среднем в 4,5 раза), однако при этом качество его переупорядочения на 30% лучше. Отметим, что результаты работы PMORSy на отдельных матрицах и классах матриц могут быть улучшены по времени и качеству при настройке параметров алгоритма улучшения разделения.

Сравнение качества и времени работы PMORSy и *mt-metis* при параметрах $nSt = 10$ и $sm = 100$ для алгоритмов разделения с 10 перезапусками приведено ниже (табл. 3).

Заключение

Переупорядочение строк и столбцов симметричных разреженных матриц — один из ключевых этапов в процессе прямого решения СЛАУ. Известно, что применяемые на данном этапе алгоритмы существенно влияют как на общее время решения задачи, так и на объем необходимой памяти. Проблема построения параллельных алгоритмов переупорядочения, сочетающих приемлемую эффективность масштабируемости с достаточным качеством получаемых перестановок, не является полностью решенной.

В данной работе изучается влияние разных алгоритмов округления и разделения графов на заполнение фактора и время работы параллельного метода вложенных сечений. В качестве экспериментальной платформы используется ранее представленная авторами параллельная библиотека PMORSy [15]. Данная библиотека реализует многоуровневый метод вложенных сечений, распараллеленный на основе логических задач OpenMP.

Одной из целей работы является повышение производительности используемого параллельного алгоритма при сохранении качества получаемых перестановок. Для этого был реализован ряд распространенных алгоритмов округления и разделения графов, изучены различные параметры алгоритмов, выполнено более 12 000 экспериментов на 48 симметричных матрицах из коллекции университета Флориды.

Эксперименты позволили сделать следующие выводы:

1. Среди алгоритмов округления наилучшие результаты по качеству и времени работы показал алгоритм паросочетания тяжелых ребер. Его использование в сравнении с другими алгоритмами позволило сократить время работы до 20% в зависимости от матрицы.
2. Рассмотренные алгоритмы разделения дают близкие по качеству результаты работы. Для большинства матриц переупорядочение выполнялось быстрее при использовании алгоритма жадного разделения или поиска в ширину из псевдопериферийной вершины. Для большинства экспериментов разница во времени работы при использовании разных алгоритмов разделения составляет около 10%.

Таблица 3

Отношение времени и качества работы PMORSy к работе mt-metis при запуске в 16 потоков. Параметры PMORSy $nSt = 10$, $sm = 100$. Наименьшие значения выделены зеленым цветом, наибольшие значения — красным цветом, промежуточные значения — желтым цветом. Результаты, дающие неприемлемое качество или время работы, отмечены белым

Качество PMORSy / mt-metis													Матрица	Время PMORSy / mt-metis												
RM LS	HEM LS	HCM LS	RM GG 10	RM GGG 10	RM KL 10	HEM GG 10	HEM GGG 10	HEM KL 10	HCM GG 10	HCM GGG 10	HCM KL 10	RM LS		HEM LS	HCM LS	RM GG 10	RM GGG 10	RM KL 10	HEM GG 10	HEM GGG 10	HEM KL 10	HCM GG 10	HCM GGG 10	HCM KL 10		
1.2	1.4	1.3	1.3	1.7	1.3			1.3	1.4	1.3	1.3	1.2	HTC_336_9129	3.3	2.6	2.5	3.1		4.0	2.5	2.6	2.8	2.6	2.7	9.3	
0.9	0.9	1.0	0.9	0.9	0.9	1.0	1.1	1.0	1.0	1.0	1.0	1.0	Lin	2.2	1.8	2.0	2.5	2.2	2.8	1.8	1.8	2.0	1.9	1.9	2.0	
1.0	1.0	1.0	1.0	1.0		1.0	1.0	1.0	1.0	1.0	1.0	1.0	boyd2	0.9	0.6	0.6	1.2	1.2	0.7	1.0	1.0	1.1	1.0	1.0	1.1	
1.4	1.4	1.4	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	darcy003	3.0	2.4	2.5	3.1	2.9	3.5	2.6	2.6	3.1	2.5	2.7	2.7	
1.0	0.9	0.8		0.9		1.7	1.1	1.3	0.9	1.6	1.9		c-big	4.8	2.9	2.5					2.6	6.7	2.8			
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	helm2d03	3.2	2.7	2.7	3.3	3.2	3.6	2.6	2.7	2.9	2.9	2.9	2.8	
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	parabolic_fem	3.7	2.7	2.8	3.7	3.5	3.7	3.0	2.7	3.1	2.9	3.0	3.2	
0.7	0.7	0.8	2.3	2.3	2.3	0.8	0.8	0.8	0.7	0.8	0.9		CO		1.0	1.0				0.9	1.0	1.0	1.0	1.2	7.6	
1.2	1.4	1.3	1.4	1.3	1.2	1.3	1.3	1.2	1.2	1.5	1.2		offshore	2.7	2.6	2.5	3.5	2.9	3.2	2.5	2.4	2.7	2.5	2.6	2.6	
0.9	0.9	0.9	1.0	1.0	0.9	0.8	1.0	0.9	1.0	1.0	1.0		apache2	3.0	2.7	2.7	3.1	3.2	3.3	2.7	2.6	2.8	2.8	2.7	2.8	
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9		ecology2	3.7	2.9	3.2	3.6	3.8	4.1	3.1	3.3	3.3	3.2	3.6	3.6	
0.8	0.9	0.9	0.8	0.9	0.8	0.9	0.9	0.9	0.9	0.9	0.9		ecology1	4.4	2.8	3.4	3.5	3.7	4.2	3.1	2.9	3.1	3.1	3.2	3.3	
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0		tmt_sym	3.7	3.2	3.3	3.7	3.8	4.0	3.1	3.1	3.2	3.2	3.2	3.3	
0.8	0.8	0.8	2.3	2.3	2.3	0.9	0.8	0.8	0.9	1.1	0.8		Si87H76	3.3	2.1	2.5				2.4	2.2	2.0	4.3	2.7	3.1	
1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1		hood	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9		BenElechi1	0.4	0.3	0.3	0.4	0.4	0.4	0.3	0.3	0.4	0.3	0.3	0.4	
1.0	1.0	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0		G3_circuit	3.4	2.9	3.0	3.4	3.5	3.5	2.7	2.9	2.9	2.9	3.1	3.2	
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0		thermal2	4.0	3.0	3.2	3.6	4.1	3.9	3.2	3.3	3.4	3.4	3.5	3.8	
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0		af_3_k101	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1.1	1.0	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1		bmw3_2	0.5	0.5	0.5	0.6	0.6	0.7	0.5	0.5	0.5	0.5	0.5	0.6	
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0		pwtk	0.4	0.4	0.4	0.5	0.5	0.5	0.4	0.4	0.4	0.4	0.4	0.5	
1.0	1.0	1.0		1.0	1.7	1.0	1.3	1.0	1.9	2.3	2.0		kkt_power	3.3	3.8	2.8	4.2	3.5		3.4	3.5	3.9				
0.7	0.7	0.7	1.0	1.7	1.4	0.7	0.7	0.7	0.7	0.7	0.7		nlpkkt80	3.9	3.1	3.2				3.0	3.0	3.2	3.2	3.2	3.4	
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0		af_shell3	0.5	0.4	0.6	0.5	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0		af_shell9	0.6	0.5	0.5	0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	
0.8	0.8	0.9	2.4	2.4	2.4	0.9	2.4	0.8	0.9	2.4	1.4		Ga41As41H72		2.6	2.6				2.6		2.7	6.9			
1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1		msdoor	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.3	0.3	0.3	0.3	0.3	
1.0	1.0	1.0	0.9		1.3	1.0	1.0	0.9	0.9	0.9	1.0		StocF-1465	3.6	2.8	3.2	8.0	4.2	4.3	3.0	3.0	3.0	3.1	3.2	3.2	
1.2	1.1	1.2	1.1	1.1	1.1	1.1	1.2	1.1	1.2	1.1	1.1		gsm_106857	1.5	1.5	1.5	1.6	1.5	1.6	1.3	1.3	1.4	1.6	1.3	1.4	
1.2	1.2	1.1	1.2	1.2	1.1	1.2	1.2	1.1	1.2	1.1	1.2		F1	1.2	1.0	0.9	1.2	1.2	1.1	0.9	1.1	1.0	1.1	1.1	1.0	
0.9	0.9	0.9	0.8	2.7	0.9	0.9	0.9	0.9	0.9	0.9	0.9		Fault_639	1.0	0.8	0.8	1.0		1.3	0.8	0.7	0.8	0.8	0.8	0.9	
1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2		inline_1	0.9	0.7	0.7	0.9	1.0	0.9	0.7	0.7	0.7	0.7	0.6	0.7	
0.9	0.9	0.9	0.8	0.8	0.9	0.9	0.9	0.8	0.9	0.9	0.9		Emilia_923	1.1	0.8	0.8	1.2	1.0	1.0	0.8	0.8	0.8	0.8	0.8	0.8	
1.1	1.2	1.1	1.2	1.3	1.3	1.2	1.1	1.1	1.2	1.1	1.2		boneS10	1.0	0.9	0.9	1.2	1.1	1.0	0.9	0.9	1.0	0.9	0.9	0.9	
1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1		ldoor	0.4	0.3	0.4	0.4	0.4	0.4	0.3	0.3	0.4	0.4	0.4	0.4	
0.9	0.9	0.9	0.9	1.0	1.7	0.9	0.9	0.9	0.9	0.9	0.9		bone010	1.1	0.8	1.0	1.2	1.2		0.8	0.8	1.0	0.8	0.9	0.8	
1.1	1.0	1.0	1.3	1.1	1.1	1.1	1.1	1.0	1.0	1.0	1.0		dielFilterV2real	1.5	1.2	1.2	1.7	1.6	1.6	1.2	1.2	1.3	1.2	1.3	1.3	
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0		af_shell10	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	
0.9	0.9	1.0	1.1	1.0	0.9	0.9	0.9	1.0	1.0	0.9	0.9		Hook_1498	1.2	1.1	1.2	1.3	1.1	1.2	1.0	1.1	0.9	1.0	1.0	1.1	
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9		Geo_1438	1.0	0.8	0.9	1.0	1.0	1.0	0.8	0.8	0.8	0.8	0.9	0.9	
1.0	0.9	0.9	1.0	0.9	0.9	1.0	1.0	1.0	1.0	1.0	0.9		Serena	1.3	0.8	0.9	1.4	1.0	1.2	0.8	1.1	0.8	1.1	0.9	1.0	
1.0	1.0	1.0	1.0	1.4	1.0	1.0	1.0	1.0	1.0	1.0	1.0		audikw_1	0.9	0.8	0.8	0.9	0.9	0.9	0.7	0.7	0.8	0.7	0.8	0.8	
1.1	1.1	1.1	1.0	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1		dielFilterV3real	0.7	0.6	0.6	0.7	0.7	0.7	0.6	0.6	0.6	0.6	0.6	0.6	
0.7	0.7	0.7	0.9	1.0	0.9	0.7	0.7	0.7	0.7	0.7	0.7		nlpkkt120	4.3	3.8	3.8				4.0	3.6	3.4	3.6	3.7	4.0	
0.8	0.8	0.8	0.8	0.9	0.9	0.8	0.8	0.8	0.8	0.8	0.8		Flan_1565	0.9	0.8	0.8	0.9	1.0	1.0	0.8	0.8	0.8	0.9	0.9	0.9	
0.7	0.7	0.7	0.9	1.2	1.1	0.7	0.7	0.7	0.7	0.7	0.7		nlpkkt160	4.1	3.5	3.8	8.7	5.9	6.7	4.3	3.4	3.6	4.1	4.0	3.9	
0.7	0.7	0.7	0.8	1.0	1.0	0.7	0.7	0.7	0.7	0.7	0.7		nlpkkt200	5.2	3.7	4.2				3.6	3.6	3.8	3.9	4.3	4.1	
0.6	0.6	0.6	0.7	0.8	0.8	0.6	0.6	0.6	0.6	0.6	0.6		nlpkkt240	5.4	4.1	4.0	9.9			3.7	4.7	4.0	4.9	4.3	5.0	

3. Работа алгоритма улучшения разделения оказывает большее влияние на качество и время переупорядочения, чем работа алгоритма разделения.
4. Для матриц с малым заполнением (до 25×10^6) целесообразно устанавливать малый порог на размер графа, к которому применяется многоуровневый алгоритм. Для матриц с большим заполнением использование малого порога приводит к существенной потере качества переупорядочения.
5. Существенным фактором, ограничивающим масштабирование текущей реализации PMORSy, является последовательная работа отдельных алгоритмов многоуровневого метода вложенных сечений на первых шагах работы метода.

В целом по результатам выполненной работы удалось выбрать комбинацию алгоритмов и их параметров, преимущественно дающую приемлемые результаты по времени и качеству получаемых перестановок. В дальнейшем планируется разработать двухуровневую схему распараллеливания, сочетающую параллелизм по задачам и внутреннее распараллеливание отдельных алгоритмов. Ожидаемый прирост производительности связан с лучшим использованием ресурсов параллелизма при решении наиболее трудоемких подзадач, возникающих в начале работы метода вложенных сечений.

Работа частично поддержана компанией Intel и Минобрнауки РФ (соглашение № 1.115.2014/К). Авторы благодарят Александра Калинин за полезные обсуждения и внимание к работе.

Литература

1. Bui T., Jones C. A Heuristic for Reducing Fill in Sparse Matrix Factorization // 6th SIAM Conference Parallel Processing for Scientific Computing. 1993. P. 445–452.
2. Chevalier C., Pellegrini F. PT-Scotch: A Tool for Efficient Parallel Graph Ordering // Parallel Computing. 2008. Vol. 34, No. 6. P. 318–331. DOI: 10.1016/j.parco.2007.12.001.
3. Davis T. A., Hu Y. The University of Florida Sparse Matrix Collection // ACM Transactions on Mathematical Software (TOMS). 2011. Vol. 38, No. 1. P. 1. DOI: 10.1145/2049662.2049663.
4. George A. et al. Sparse Cholesky Factorization on a Local-memory Multiprocessor // SIAM J. on Scientific and Statistical Computing. 1988. Vol. 9, No. 2. P. 327–340. DOI: 10.1016/0167-8191(92)90014-x.
5. George A. Nested Dissection of a Regular Finite Element Mesh // SIAM J. on Numerical Analysis. 1973. Vol. 10, No. 2. P. 345–363. DOI: 10.1137/0710032.
6. George A., Liu J.W.H. An Automatic Nested Dissection Algorithm for Irregular Finite Element Problems // SIAM J. on Numerical Analysis. 1978. Vol. 15, No. 5. P. 1053–1069. DOI: 10.1137/0715069.
7. Karypis G., Kumar V. ParMetis: Parallel Graph Partitioning and Sparse Matrix Ordering Library. Tech. Rep. TR 97-060, University Of Minnesota, Department Of Computer Science. 1997.
8. Karypis G., Kumar V. A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs // SIAM J. on Scientific Computing. 1998. Vol. 20, No. 1. P. 359–392. DOI: 10.1137/s1064827595287997.
9. Karypis G., Kumar V. Analysis of Multilevel Graph Partitioning // Proceedings of the 1995 ACM/IEEE conference on Supercomputing. ACM, 1995. P. 29.

10. Karypis G., Kumar V. METIS. A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-reducing Orderings of Sparse Matrices. Technical Report University of Minnesota, Department of Computer Science and Engineering. 1998.
11. LaSalle D., Karypis G. Efficient Nested Dissection for Multicore Architectures // EuroPar 2015: Parallel Processing. 2015, Springer Berlin Heidelberg. P. 467–478.
12. Pellegrini F. Scotch and libScotch 6.0 User's Guide. Technical Report LaBRI, 2012.
13. Pellegrini F. Shared Memory Parallel Algorithms in Scotch 6 // MUMPS User Group Meeting. 2013. URL: http://graal.ens-lyon.fr/mumps/doc/ud_2013/pellegrini.pdf (дата обращения: 01.12.2016)
14. Pirova A., Meyerov I. MORsY — a New Tool for Sparse Matrix Reordering // An International Conference on Engineering and Applied Sciences Optimization M. Papadrakakis, M.G. Karlaftis, N.D. Lagaros (eds.) (Kos Island, Greece, 4-6 June 2014). P. 1952–1964.
15. Pirova A., Meyerov I., Kozinov E., Lebedev S. PMORsY: Parallel Sparse Matrix Ordering Software for Fill-in Minimization // Optimization Methods and Software. 2017. Vol. 32, No. 2. P. 274–289. DOI: 10.1080/10556788.2016.1193177.
16. Pothén A. Graph Partitioning Algorithms with Applications to Scientific Computing // Parallel Numerical Algorithms. Springer Netherlands, 1997. P. 323–368. DOI: 10.1007/978-94-011-5412-3_12.
17. Tinney W., Walker J. Direct Solutions of Sparse Network Equations by Optimally Ordered Triangular Factorization // Proceedings of the IEEE. 1967. Vol. 55, No. 11. P. 1801–1809. DOI: 10.1109/proc.1967.6011.
18. Yannakakis M. Computing the Minimum Fill-in is NP-complete // SIAM J. on Algebraic and Discrete Methods. 1981. Vol. 2, No. 1. P. 77–79. DOI: 10.1137/0602010.
19. Пирова А.Ю., Мееров И.Б., Козинев Е.А., Лебедев С.А. Параллельный алгоритм многоуровневого метода вложенных сечений для вычислительных систем с общей памятью // Вычислительные методы и программирование. 2015. Т. 16. № 3. С. 407–420.
20. Старостин Н.В. Многоуровневый итерационный алгоритм декомпозиции графа. // Системы управления и информационные технологии. 2015. Т. 61. № 3. С. 27–30.

Пирова Анна Юрьевна, ассистент, кафедра математического обеспечения и суперкомпьютерных технологий, Институт информационных технологий, математики и механики, Нижегородский государственный университет им. Н.И. Лобачевского (Нижний Новгород, Российская Федерация)

Кудрявцев Никита Юрьевич, магистрант, кафедра математического обеспечения и суперкомпьютерных технологий, Институт информационных технологий, математики и механики, Нижегородский государственный университет им. Н.И. Лобачевского (Нижний Новгород, Российская Федерация)

Мееров Иосиф Борисович, к.т.н., доцент, зам. заведующего кафедрой, кафедра математического обеспечения и суперкомпьютерных технологий, Институт информационных технологий, математики и механики, Нижегородский государственный университет им. Н.И. Лобачевского (Нижний Новгород, Российская Федерация)

**EXPERIMENTAL EVALUATION OF ALGORITHMS
IN THE PARALLEL MULTILEVEL NESTED DISSECTION METHOD**

© 2017 A.Yu. Pirova, N.Yu. Kudriavtsev, I.B. Meyerov

*Lobachevsky State University of Nizhni Novgorod**(23 Gagarin Avenue, Nizhni Novgorod, 603022 Russia)**E-mail: anna.pirova@itmm.unn.ru, n.yu.kudriavtsev@gmail.com,**iosif.meyerov@itmm.unn.ru*

Received: 20.10.2016

Direct methods for solving large sparse systems of linear equations make use of reordering of rows and columns of the original matrix. The goal of this procedure is to reduce the fill-in during the subsequent numerical factorization. Finding the ordering with the minimum fill-in is NP-complete. Heuristic methods are used to solve this problem. These methods can be evaluated for both quality (fill-in) and time to obtain an ordering. The multilevel nested dissection method performs reasonably well in terms of both criteria and is one of the most widely used reordering methods. The method has some parallelization potential, which is utilized in several implementations (ParMETIS, mtMETIS, PT-SCOTCH, PMORSy). However, low arithmetic intensity, irregular memory access pattern, workload imbalance and the trade-off between run time and quality motivates further investigation of the method.

This paper presents the comparison of the algorithms used on several stages of the multilevel nested dissection method in terms of fill-in and run time on a parallel system. The implementation and experiments are done using the parallel PMORSy library, which outperforms competitors on some matrices from the University of Florida sparse matrix collection. As the result we distinguish the most promising combination of the algorithms and improve the quality and performance of PMORSy.

Keywords: multilevel nested dissection, sparse matrix ordering, parallel algorithm.

FOR CITATION

Pirova A.Yu., Kudriavtsev N.Yu., Meyerov I.B. Experimental Evaluation of Algorithms in the Parallel Multilevel Nested Dissection Method. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2017. vol. 6, no. 1. pp. 38–55. (in Russian) DOI: 10.14529/cmse170103.

References

1. Bui T., Jones C. A Heuristic for Reducing Fill in Sparse Matrix Factorization. *6th SIAM Conference Parallel Processing for Scientific Computing*. 1993. pp. 445–452.
2. Chevalier C., Pellegrini F. PT-Scotch: A Tool for Efficient Parallel Graph Ordering. *Parallel Computing*. 2008. vol. 34, no. 6. pp. 318–331. DOI: 10.1016/j.parco.2007.12.001.
3. Davis T. A., Hu Y. The University of Florida Sparse Matrix Collection. *ACM Transactions on Mathematical Software (TOMS)*. 2011. vol. 38, no. 1. pp. 1. DOI: 10.1145/2049662.2049663.
4. George A. et al. Sparse Cholesky Factorization on a Local-memory Multiprocessor. *SIAM J. on Scientific and Statistical Computing*. 1988. vol. 9, no. 2. pp. 327–340. DOI: 10.1016/0167-8191(92)90014-x.
5. George A. Nested Dissection of a Regular Finite Element Mesh. *SIAM J. on Numerical Analysis*. 1973. vol. 10, no. 2. pp. 345–363. DOI: 10.1137/0710032.

6. George A., Liu J.W.H. An Automatic Nested Dissection Algorithm for Irregular Finite Element Problems. *SIAM J. on Numerical Analysis*. 1978. vol. 15, no. 5. pp. 1053–1069. DOI: 10.1137/0715069.
7. Karypis G., Kumar V. *ParMetis: Parallel Graph Partitioning and Sparse Matrix Ordering Library*. Tech. Rep. TR 97-060, University Of Minnesota, Department Of Computer Science. 1997.
8. Karypis G., Kumar V. A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs. *SIAM J. on Scientific Computing*. 1999. vol. 20, no. 1. pp. 359–392. DOI: 10.1137/s1064827595287997.
9. Karypis G., Kumar V. Analysis of Multilevel Graph Partitioning. *Proceedings of the 1995 ACM/IEEE conference on Supercomputing*. ACM, 1995. pp. 29.
10. Karypis G., Kumar V. *METIS. A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-reducing Orderings of Sparse Matrices*. University of Minnesota, Department of Computer Science and Engineering. 1998.
11. LaSalle D., Karypis G. Efficient Nested Dissection for Multicore Architectures. *Euro-Par 2015: Parallel Processing*. 2015, Springer Berlin Heidelberg. pp. 467–478.
12. Pellegrini F. *Scotch and libScotch 6.0 User's Guide*. Technical Report LaBRI, 2012.
13. Pellegrini F. Shared Memory Parallel Algorithms in Scotch 6. *MUMPS User Group Meeting*. 2013. Available at: http://graal.ens-lyon.fr/mumps/doc/ud_2013/pellegrini.pdf (accessed: 01.12.2016)
14. Pirova A., Meyerov I. MORSy — a New Tool for Sparse Matrix Reordering. *An International Conference on Engineering and Applied Sciences Optimization*. M. Papadrakakis, M.G. Karlaftis, N.D. Lagaros (eds.) Kos Island, Greece, 4–6 June 2014. pp. 1952–1964.
15. Pirova A., Meyerov I., Kozinov E., Lebedev S. PMORSy: Parallel Sparse Matrix Ordering Software for Fill-in Minimization. *Optimization Methods and Software*. 2017. vol. 32, no. 2, pp. 274–289. DOI: 10.1080/10556788.2016.1193177.
16. Pothen A. Graph Partitioning Algorithms with Applications to Scientific Computing. *Parallel Numerical Algorithms*. Springer Netherlands, 1997. pp. 323–368. DOI: 10.1007/978-94-011-5412-3_12.
17. Tinney W., Walker J. Direct Solutions of Sparse Network Equations by Optimally Ordered Triangular Factorization. *Proceedings of the IEEE*. 1967. vol. 55, no. 11. pp. 1801–1809. DOI: 10.1109/proc.1967.6011.
18. Yannakakis M. Computing the Minimum Fill-in is NP-complete. *SIAM J. on Algebraic and Discrete Methods*. 1981. vol. 2, no. 1. pp. 77–79. DOI: 10.1137/0602010.
19. Pirova A.Yu., Meyerov I.B., Kozinov E.A., Lebedev S.A. A Parallel Multilevel Nested Dissection Algorithm for Shared-memory Computing Systems. *Numerical Methods and Programming*. 2015. vol. 16, no. 3. pp. 407–420. (in Russian)
20. Starostin N.V. The Multilevel Iteration Algorithm for Graph Decomposition. *Sistemy upravleniya i informacionnye tehnologii* [Operation Systems and Information Technologies]. 2015. vol. 61, no. 3. pp. 27–30. (in Russian)

ИССЛЕДОВАНИЕ ВЛИЯНИЯ СТАЦИОНАРНЫХ ТЕЧЕНИЙ НА ДИНАМИЧЕСКИЕ ПРОЦЕССЫ И ЭВОЛЮЦИЮ ЗАГРЯЗНЕНИЙ В АЗОВСКОМ МОРЕ

© 2017 г. **Л.В. Черкесов**, **Т.Я. Шульга**

Федеральное государственное бюджетное учреждение науки «Морской гидрофизический институт РАН» (299011, Севастополь, ул. Капитанская, д. 2)

E-mail: tyshulga@mail.ru

Поступила в редакцию: 20.10.2016

С использованием трехмерной нелинейной математической модели изучаются динамические процессы, и особенности трансформации примеси в Азовском море, вызванные действием переменного ветра и атмосферного давления при наличии фоновых стационарных течений. На основании результатов численных расчетов сделаны выводы о влиянии скоростей стационарных течений на максимальные отклонения уровня и скорости нестационарных течений, генерируемых полями ветра и атмосферного давления, полученными по данным модели SKIRON. Выполнен анализ влияния изменения интенсивности стационарных течений на размеры областей осушения и затопления в прибрежных районах моря в зависимости от угла наклона (подъема) рельефа береговой зоны. Показано, что совместное действие постоянного ветра с полями прогностической модели SKIRON приводит к существенному увеличению площади распространения пассивной примеси и времени ее рассеивания по сравнению с воздействием только стационарных течений.

Ключевые слова: численное моделирование, сигма-координатная модель, динамические процессы, стационарные течения, области осушения и затопления, эволюция пассивной примеси.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Черкесов Л.В., Шульга Т.Я. Исследование влияния стационарных течений на динамические процессы и эволюцию загрязнений в Азовском море // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. Т. 6, № 1. С. 56–72. DOI: 10.14529/cmse170104

Введение

Последствия интенсивной промышленной и сельскохозяйственной деятельности в Азово-Черноморском регионе оказывают заметное негативное влияние на морскую среду. В связи с этим возникает необходимость проведения научных исследований с целью оценить последствия этого влияния и найти пути его существенного уменьшения. Одним из направлений таких исследований является анализ сгонно-нагонных процессов и прогнозирование размеров районов возможных затоплений и осушений в прибрежных областях Азовского моря.

Изучение течений и колебаний уровня, возникающих в Азовском море под действием переменного во времени ветра и атмосферного давления при отсутствии стационарных течений, проведено в [3, 4, 8]. На основе математического моделирования и данных наблюдений анализируется механизм формирования экстремальных течений и кратковременных наводнений, наблюдаемых часто в отдельных районах побережья Азовского моря, приводящих иногда к человеческим жертвам и материальному ущербу. Выделены районы побережья, в которых может действовать подобный механизм формирования наводнений и экстремальных течений [2, 5, 6]. Из современных публикаций по моделиро-

ванию циркуляции выделяются численные исследования, проведенные в [2, 10], в которых применялись трехмерные нелинейные модели мелкой воды для изучения стационарных движений и переноса примесей. С помощью двухслойной математической модели и по материалам ежедневных гидрометеорологических наблюдений на береговой базе Южного научного центра РАН в период с 20 марта по 26 марта 2013 года, проведено восстановление картины аномального затопления дельты Дона [2].

В настоящей работе с использованием трехмерной нелинейной сигма-координатной модели РОМ (*Princeton Ocean Model*) [11] проведен анализ, объединяющих четыре этапа исследований, посвященных изучению влияния интенсивности фоновых стационарных течений в Азовском море в условиях реально действующего нестационарного ветра. В рамках цикла численных экспериментов при наличии и отсутствии фоновых течений изучены сгонно-нагонные явления и течения в море, вызываемые действием ветра, приповерхностные поля которого получены по данным реанализа модели SKIRON [13]. В результате отдельного этапа исследований, определены зависимости размеров территорий возможных осушений и затоплений, от скорости действующего постоянного ветра, и геометрии побережья. Одной из основных частей решения задачи моделирования динамических процессов в Азовском море, явился этап сравнения результатов численных расчетов с данными натурных наблюдений, полученных во время действия прогностических полей, на ряде гидрологических станций. Анализ результатов заключительного цикла численных исследований позволяет сделать вывод о зависимости характеристик эволюции областей загрязнения на различных горизонтах Азовского моря от наличия прогностических полей и скоростей фоновых стационарных течений.

Статья организована следующим образом. В разделе 1 рассмотрено общее описание уравнений используемой в расчетах математической модели, переход к криволинейным координатам, дано обоснование ее параметров, выбранных с учетом различных физических факторов, объяснены особенности ее численной реализации. Раздел 2 содержит информацию о полях ветра и атмосферного давления, используемых в вычислительных экспериментах. В разделе 3 проводится обсуждение результатов расчетов моделирования динамических процессов в Азовском море, исследованию волн и течений, генерируемых неоднородными по времени и пространству полями ветра и атмосферного давления при наличии и отсутствии фоновых стационарных течений. В заключении подводятся итоги работы и формулируются направления дальнейших исследований.

1. Постановка задачи. Граничные и начальные условия

Поставим своей задачей исследовать движение жидкости, вызываемое атмосферными возмущениями, действующими на поверхность Азовского моря. Исходными уравнениями используемой математической модели является система, описывающая движение однородной вязкой несжимаемой жидкости в приближении теории мелкой воды [7, 9, 11]. Введем систему координат, в которой ось x направлена на восток, y — на север, z — вертикально вверх

$$\frac{du}{dt} - fv + \frac{1}{\rho} \frac{\partial P}{\partial x} = \frac{\partial \tau_{11}}{\partial x} + \frac{\partial \tau_{12}}{\partial y} + \frac{\partial}{\partial z} K_M \frac{\partial u}{\partial z}, \quad (1)$$

$$\frac{dv}{dt} + fu + \frac{1}{\rho} \frac{\partial P}{\partial y} = \frac{\partial \tau_{21}}{\partial x} + \frac{\partial \tau_{22}}{\partial y} + \frac{\partial}{\partial z} K_M \frac{\partial v}{\partial z}, \quad (2)$$

$$\frac{\partial P}{\partial z} + g\rho = 0, \quad (3)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \quad (4)$$

При этом u, v, w — проекции скорости по осям x, y, z ; t — время; g — ускорение свободного падения; f — параметр Кориолиса, $\zeta(x, y, t)$ — профиль свободной поверхности, $d/dt = \partial/\partial t + u\partial/\partial x + v\partial/\partial y + w\partial/\partial z$ — полная производная; $P(x, y, t) = P_{atm} + \rho g(\zeta - z)$ — давление, где $P_{atm} = 1013,25$ гПа — стандартное атмосферное давление при температуре 0° на широте 45° ; ρ — плотность воды, $\rho_0 = (\zeta + H)^{-1} \int_{-H}^{\zeta} \rho dz$ — средняя плотность воды по глубине. Параметризация коэффициентов вертикальной вязкости K_M и турбулентной диффузии K_H выполняются в соответствии с полуэмпирической дифференциальной моделью Меллора—Ямады [14]. Коэффициент горизонтальной вязкости A_M вычисляется с использованием модели подсеточной вязкости в зависимости от горизонтальных градиентов скорости [15]. Компоненты тензора турбулентных напряжений Рейнольдса τ_{ij} в (1), (2) определяются через градиенты скорости средних течений согласно формулам

$$\tau_{11} = 2A_M \frac{\partial u}{\partial x}, \tau_{12} = \tau_{21} = A_M \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right), \tau_{22} = 2A_M \frac{\partial v}{\partial y}. \quad (5)$$

На свободной поверхности удовлетворяются кинематическое и динамические условия.

$$w|_{z=\zeta} = \frac{\partial \zeta}{\partial t} + u \frac{\partial \zeta}{\partial x} + v \frac{\partial \zeta}{\partial y} + w \frac{\partial \zeta}{\partial z}, \quad K_M \left(\frac{\partial u}{\partial z}, \frac{\partial v}{\partial z} \right) \Big|_{z=\zeta} = (\tau_{0x}, \tau_{0y}), \quad (6)$$

при этом $\tau_{0x} = C_a W_x |\mathbf{W}|$ и $\tau_{0y} = C_a W_y |\mathbf{W}|$ — проекции касательных напряжений ветра [11]; \mathbf{W} — вектор скорости ветра на высоте 10 м над уровнем моря; C_a — эмпирический коэффициент поверхностного трения [16], который варьируется в зависимости от величины скорости ветра

$$10^3 C_a = \begin{cases} 2,5; & |\mathbf{W}| > 22 \text{ м/с} \\ 0,49 + 0,065|\mathbf{W}|; & 8 \leq |\mathbf{W}| \leq 22 \text{ м/с} \\ 1,2; & 4 \leq |\mathbf{W}| \leq 8 \text{ м/с} \\ 1,1; & 1 \leq |\mathbf{W}| \leq 4 \text{ м/с} \end{cases}. \quad (7)$$

На дне ($z = -H(x, y)$) нормальная составляющая скорости равна нулю, придонные касательные напряжения связаны со скоростью квадратичной зависимостью [3, 4].

$$\left(w + u \frac{\partial H}{\partial x} + v \frac{\partial H}{\partial y} \right) \Big|_{z=-H} = 0, \quad K_M \left(\frac{\partial u}{\partial z}, \frac{\partial v}{\partial z} \right) \Big|_{z=-H} = (\tau_{1x}, \tau_{1y}), \quad (8)$$

де $\tau_{1x} = C_b u \sqrt{u^2 + v^2}$, $\tau_{1y} = C_b v \sqrt{u^2 + v^2}$, C_b — коэффициент донного трения, который находится по формуле $C_b = \max\{k^2(\ln h_b/z_0)^{-2}; 0,0025\}$, где h_b — шаг по вертикали в придонном слое; $z_0 = 0,003$ м — параметр шероховатости, характеризующий гидродинамические свойства подстилающей донной поверхности. Значения z_0 определяются с помощью теории Гранта—Мадсена, описывающей механизм влияния волн на течения в придонном пограничном слое. На твердых боковых границах выполняются условия прилипания. В начальный момент времени ($t = 0$) движение жидкости отсутствует, свободная поверхность горизонтальна.

Для расчета распространения примеси концентрацией $C(x, y, z, t)$ используем уравнение переноса и диффузии [3, 5]

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left(A_H \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y} \left(A_H \frac{\partial C}{\partial y} \right) + \frac{\partial}{\partial z} \left(K_H \frac{\partial C}{\partial z} \right). \quad (9)$$

Здесь A_H — коэффициент горизонтальной и K_H — вертикальной турбулентной диффузии. К динамическим граничным условиям на свободной поверхности и в придонном слое (6) — (9) добавляются условия отсутствия потоков примеси через свободную поверхность, боковые стенки (S) и дно бассейна.

$$\left(K_H \frac{\partial c}{\partial n}\right)\Big|_{z=\zeta} = 0, \quad \left(A_H \frac{\partial c}{\partial n}\right)\Big|_S = 0, \quad \left(K_H \frac{\partial c}{\partial n}\right)\Big|_{z=-H} = 0. \quad (10)$$

При этом рассматривается цилиндрическая область загрязнения заданного радиуса, расположенная в поверхностном слое центральной части моря, глубиной равной толщине поверхностного слоя. В начальный момент времени ($t = t_0$) распределение примеси задается соотношением

$$C_0(x, y, z, t_0) = \begin{cases} 1; & r < R, 0 \geq z \geq h_1; \\ 0; & r \geq R, z < 0; \quad r < R, z < h_1, \end{cases} \quad (11)$$

где h_1 — толщина этого слоя; R — радиус области загрязнения, $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ — расстояние от центра (x_0, y_0) до точки, в которой оно вычисляется. В качестве параметров, характеризующих эволюцию пассивной примеси, выбраны: время ее рассеивания (t_d), коэффициент максимальной площади ее распространения на различных горизонтах (K_{\max}) и соответствующий этому момент времени (t_{\max}). Соотношение для расчета площади охваченной загрязнением имеет вид: $K_{\max} = S_{\max}/S_0$, где S_0 — площадь области начального загрязнения в поверхностном слое, S_{\max} — наибольшее значение этой площади на рассматриваемом горизонте в процессе трансформации примеси. Условием полного рассеивания загрязнения принимается величина ее концентрации, не превышающая $2,5 \cdot 10^{-2}$ во всей акватории моря.

Для построения численного алгоритма использована модель РОМ [11], адаптированная к условиям бассейна Азовского моря [3, 7, 8]. Выбор шагов интегрирования по временным и пространственным координатам осуществляется в соответствии с критерием устойчивости для баротропных волн [12]. Пространственное разрешение модели по широте и долготе составляет $(1/59) \times (1/84)$, при котором линейные размеры ячейки $\Delta x = \Delta y = 1,4$ км, количество узлов горизонтальной сетки равно 276×176 . Число расчетных уровней по вертикали — одиннадцать. Уравнения интегрируются с шагом $\Delta t = 18$ с для определения осредненных двумерных компонент скорости и уровня моря и $\Delta t_A = 10 \Delta t$ (3 мин) — для вычисления отклонений от найденных средних и вертикальной компоненты скорости. Для численной реализации поставленной задачи выполняется переход от координаты z к сигма-координате [3, 7, 11].

Топография дна Азовского моря на модельную сетку интерполирована с использованием массива глубин, приведенного в навигационных картах. Отклонения уровня моря анализируются на девяти станциях, расположенных вблизи крупных населенных пунктов.

2. Информация о полях ветра и атмосферного давления, используемых в вычислительных экспериментах

Модельная система SKIRON создавалась и развивалась в Афинском Университете группой Атмосферного Моделирования и Прогноза Погоды (Atmospheric Modeling and Weather Forecasting Group) [13]. Модель базируется на ETA модели, которая первоначально разрабатывалась в Университете Белграда. Основное развитие модели ETA было обеспечено NCEP. Результаты прогноза по модели SKIRON, используемые в данной ра-

боте, были получены МГИ РАН в качестве полноправного участника проекта MFSTEP. Данный вариант модели дает детальный 72-часовой прогноз метеорологических параметров для Азово-Черноморского и Средиземноморского бассейнов. Первые 48 часов вывод данных осуществляется через 2 часа, далее значения выводятся через 6 часов. Расчет параметров производится на сетке с шагом 0,1 градус широты по осям x и y . Всего выводится 16 различных параметров, в том числе, данные по скорости приводного ветра. Данные модели SKIRON были интерполированы на расчетную сетку бассейна Азовского моря с горизонтальным разрешением 1,4 км.

В атмосфере наблюдается большое разнообразие атмосферных движений, что обусловлено влиянием сил различного происхождения. Макромасштабные атмосферные процессы большой длительности с масштабом порядка тысяч километров (планетарный масштаб, зональные переносы), учет которых необходим при долгосрочных прогнозах динамики вод природных бассейнов, образуют единую систему над всем земным шаром и являются колебаниями глобальной циркуляции атмосферы. Для краткосрочных прогнозов особый интерес представляют мезомасштабные атмосферные движения, имеющие горизонтальные размеры порядка 100 км и микромасштабные с масштабом движения порядка 10 км и менее, связанные с местными влияниями топографии в самом ограниченном масштабе и с мелкомасштабными вихрями.

В данной работе исследовано влияние динамики вод Азовского моря, вызванной системами движения атмосферы синоптического масштаба, имеющими указанные горизонтальные размеры и время существования до нескольких суток на процессы, возникающие при микромасштабной атмосферной циркуляции. Исходным материалом атмосферной циркуляции явились массивы данных региональной атмосферной модели SKIRON для типичного, сезонного случая штормовой погоды с 8 по 18 сентября 2007 г. В результате серии численных экспериментов сделаны выводы о влиянии стационарных течений на динамические процессы и эволюцию загрязнений в Азовском море, вызванные неоднородными по времени и пространству атмосферными полями ветра и давления.

Стационарные течения в Азовском море вызываются действием ветра, скорость которого на поверхности моря не зависит от x и y , а первые три часа ($0 < t < 3$ ч) нарастает со временем по линейному закону, достигает наибольшего значения и далее не меняется ($t \geq 3$ ч). При этом направления однородных по пространству и времени ветровых воздействий отвечают характерным для Азово-Черноморского региона [1, 6] зональным восточному и западному ветрам.

Момент выхода течений на установившийся режим ($t = t_0$) определяется тем, что между двумя соседними значениями времени ($t_k > t_0$ и $t_k + \Delta t$) не происходит заметных изменений отклонений уровня и скоростей течений (изменения не превышают 3%). Исходя из этого, на основании серии численных экспериментов, находим время установления стационарных фоновых течений ($t_0=48$ ч).

Кроме постоянного однородного ветра акватория Азовского моря находится под воздействием полей переменного по времени и неоднородного по пространству ветра и атмосферного давления. В качестве таких полей используются поля приводного ветра атмосферной модели SKIRON [13], отвечающие штормовой ситуации с 8 по 18 сентября 2007 г. Действие этого (прогностического) ветра при наличии и отсутствии фоновых течений в море численно исследуется в течение 10 сут. Результаты расчетов первых трех суток не

анализируются и применяются для получения начальных данных о скоростях течений и уровне моря на 0 ч 11 сентября 2007 г.

От момента установления движения жидкости ($t > t_0$) к постоянному ветру ($\mathbf{W}_{st}=\text{const}$, $t \geq 3$ ч) присоединяется действие неоднородного поля ветра, полученного по данным атмосферной модели SKIRON (\mathbf{W}_{SKIRON}). При этом, постоянное давление (P_{atm}) в (1) — (4) заменяется переменным значением P_{SKIRON} . Таким образом, от момента времени $t > t_0$ квазистационарную составляющую течений в море поддерживает \mathbf{W}_{st} , а нестационарную добавляет \mathbf{W}_{SKIRON} . Моделирование фоновых течений выполняется западным ветром со скоростями $\mathbf{W}_{st}^1=5$ м/с и $\mathbf{W}_{st}^2=10$ м/с.

В табл. 1 приведены максимальные скорости при различном направлении ветра \mathbf{W}_{SKIRON} как функции времени (над Азовским морем). При этом между двумя соседними значениями времени скорость ветра меняется монотонно. Величина отклонения вектора скорости ветра от направления оси x (на восток под углом 50° к параллели) указана в градусах. Отсюда видно, что за рассматриваемый период (192 ч) наибольшее, среднее и наименьшее значения максимальной скорости равны 12,7; 6,7 и 1,6 м/с соответственно. Преобладающими направлениями ветра являются северо-восточное и северо-западное.

Таблица 1

Максимальные скорости ветра как функции времени, полученные по данным модели SKIRON с 0 ч 11 сентября до 24 ч 18 сентября 2007 г.

Время, ч	Скорость ветра, м/с	Направление ветра, град	Время, ч	Скорость ветра, м/с	Направление ветра, град
2	6,8	107	76	9,4	354
14	2,6	344	92	12,7	350
28	5,8	9	104	9,6	100
30	3,0	10	106	7,5	210
32	5,2	107	108	5,7	200
44	5,8	344	112	3,5	110
48	8,1	100	124	5,2	344
50	4,2	213	130	3,3	354
54	5,8	195	132	1,6	347

На рис. 1 показаны векторные поля прогностического ветра над Азовским морем, которые соответствуют экстремальным значениям его скорости или резким переменам направления. Первая экстремальная ситуация возникает через 36 ч (рис. 1, *a*). Над южной частью отмечается антициклоническое воздействие, а в северной части моря скорость ветра на отдельных участках достигает 10 м/с. Через 68 ч со скоростью 15 — 20 м/с перемещаются два небольших циклона (рис. 1, *b*), которые формируются в восточной и западной частях моря. Следующие максимумы скоростей ветра приходятся на 90 и 140 ч, они соответствуют обширным циклоническим образованиям (рис. 1, *в*, *г*). На рис. 1, *д*, *e* показаны поля ветра с экстремумами скорости в моменты времени 152 и 168 ч. В дальнейшем до 18 сентября над Азовским морем скорость ветра уменьшается и его направление практически не меняется.

Как следует из анализа величин атмосферного давления P_{SKIRON} в период с 8 по 18 сентября 2007 г., штормовая ситуация сопровождалась существенным понижением атмосферного давления при прохождении циклона 16 сентября.

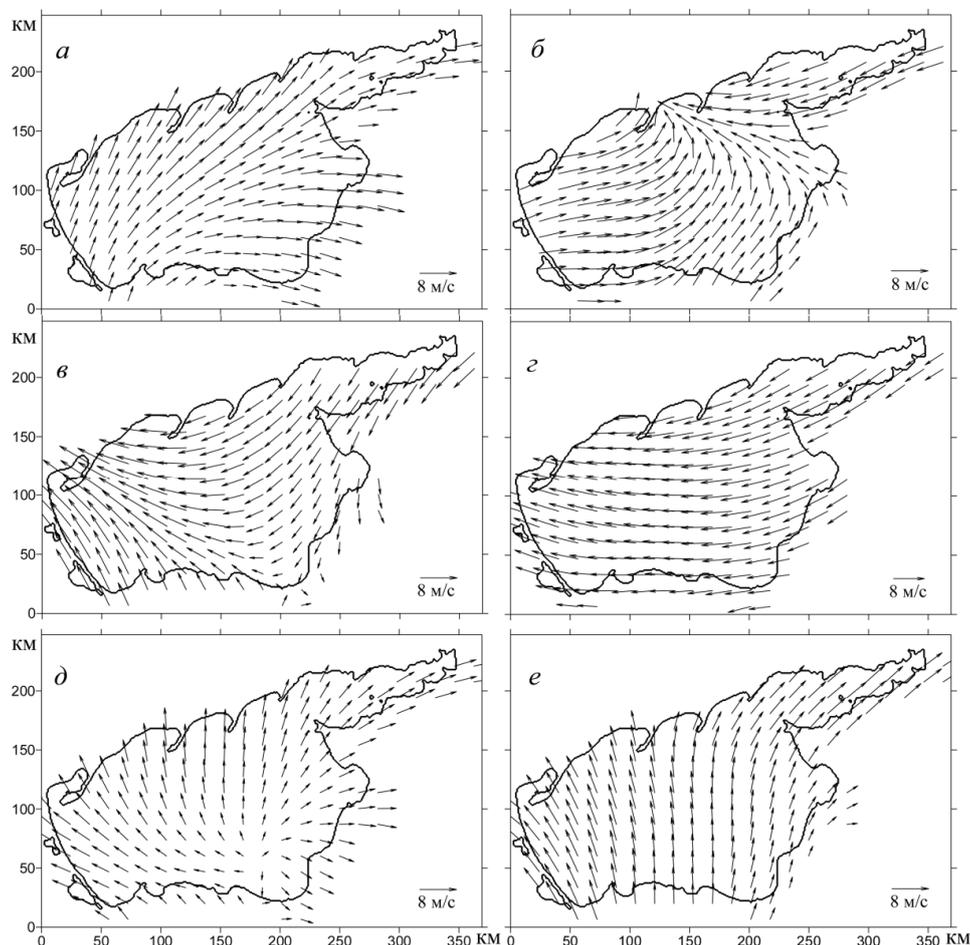


Рис. 1. Поля прогностического ветра в различные моменты времени от 0 ч 11 сентября 2007 г.: 36 ч (а); 68 ч (б); 90 ч (в); 140 ч (г); 152 ч (д); 168 ч (е)

3. Анализ результатов численных расчетов

3.1. Исследование влияния фоновых течений на динамические процессы в Азовском море, возникающие под действием атмосферных полей, полученных по данным модели SKIRON

В первой серии численных экспериментов исследуется влияние совместного действия ветров ($\mathbf{W}_{st} + \mathbf{W}_{SKIRON}$ — стационарного и прогностического) на максимальные скорости течений и экстремальные значения сгонов и нагонов. Полученные результаты сравниваются с данными расчетов, выполненных при моделировании волновых процессов в Азовском море, которые вызваны только постоянным или только прогностическим ветром. Численные эксперименты реализованы для двух скоростей западного стационарного ветра с целью исследования влияния возникающих при этом фоновых

течений на отклонения уровня моря и поля скоростей неустановившихся течений, вызванных в дальнейшем действием полей прогностического ветра.

В табл. 2 приведены максимальные значения скоростей течений, вызванных постоянным ($\mathbf{W}_{st}^{1,2}$) и только прогностическим (\mathbf{W}_{SKIRON}) ветрами, а также их совместным действием ($\mathbf{W}_{st}^{1,2} + \mathbf{W}_{SKIRON}$).

Таблица 2

Максимальные значения скоростей течений ($|\mathbf{U}|_{\max}$) на различных глубинах Азовского моря, время их достижения (t_{\max}) и соответствующие им координаты

Скорость ветра	Горизонт, м	$ \mathbf{U} _{\max}$, м/с	x_{\max} , км	y_{\max} , км	t_{\max} , ч. мин
$\mathbf{W}_{st}^1 = 5$ м/с	1	0,16	236,29	174,08	8.00
	3	0,13	236,39	173,44	8.02
	5	0,12	237,01	173,31	8.22
	10	0,08	235,34	173,00	9.44
$\mathbf{W}_{st}^2 = 10$ м/с	1	0,49	235,38	172,19	8.35
	3	0,44	235,97	172,86	8.55
	5	0,41	237,07	171,50	8.26
	10	0,29	235,99	171,16	9.34
\mathbf{W}_{SKIRON}	1	0,48	206,25	189,27	128
	3	0,46	241,76	188,88	130
	5	0,41	263,30	199,05	137
	10	0,35	238,20	175,76	145
$\mathbf{W}_{st}^1 + \mathbf{W}_{SKIRON}$ N	1	0,73	201,48	106,15	128
	3	0,69	224,73	176,44	130
	5	0,61	225,31	156,78	137
	10	0,45	174,86	160,92	145
$\mathbf{W}_{st}^2 + \mathbf{W}_{SKIRON}$	1	0,91	159,09	175,58	152
	3	0,85	199,38	159,67	154
	5	0,72	258,19	188,40	157
	10	0,65	209,66	146,44	162

Из анализа этих данных следует, что при совместном действии ветров максимумы скоростей нестационарных течений увеличиваются. Так, при $\mathbf{W}_{st}^1 + \mathbf{W}_{SKIRON}$ величина $|\mathbf{U}|_{\max}$ в 1,52; 1,50; 1,48 и 1,29 раза больше, чем при \mathbf{W}_{SKIRON} на горизонтах 1; 3; 5 и 10 м соответственно. Аналогично при $\mathbf{W}_{st}^2 + \mathbf{W}_{SKIRON}$ значение $|\mathbf{U}|_{\max}$ в 1,9; 1,84; 1,75 и 1,58 раза больше, чем при \mathbf{W}_{SKIRON} на тех же горизонтах. Легко видеть, что для всех значений скоростей ветра \mathbf{U}_{\max} монотонно убывает с увеличением горизонта. Отметим также, что значения \mathbf{U}_{\max} на рассматриваемых горизонтах при \mathbf{W}_{st}^1 и \mathbf{W}_{SKIRON} отличаются не более чем на 5%.

В табл. 3 приведены максимальные повышения (ζ_{\max}) и понижения ($|\zeta|_{\max}$) уровня при $\mathbf{W}_{\text{st}}^{1,2}$, $\mathbf{W}_{\text{SKIRON}}$ и $\mathbf{W}_{\text{st}}^{1,2} + \mathbf{W}_{\text{SKIRON}}$. Видно, что при \mathbf{W}_{st}^2 величины ζ_{\max} и $|\zeta|_{\max}$ в 3,45 и 3,9 раза больше, чем при \mathbf{W}_{st}^1 . При $\mathbf{W}_{\text{SKIRON}}$ значения ζ_{\max} и $|\zeta|_{\max}$ в 2,15 и 2,5 раза больше, чем при \mathbf{W}_{st}^1 , и в 1,6 и 1,5 раза меньше, чем при \mathbf{W}_{st}^2 . При $\mathbf{W}_{\text{st}}^2 + \mathbf{W}_{\text{SKIRON}}$ величина ζ_{\max} в 1,31, а $|\zeta|_{\max}$ в 2,1 раза больше, чем при $\mathbf{W}_{\text{st}}^1 + \mathbf{W}_{\text{SKIRON}}$. Таким образом, действие стационарного ветра одного направления, но большей скорости приводит при наличии $\mathbf{W}_{\text{SKIRON}}$ к возрастанию наибольших отклонений уровня на 31 и 107% для ζ_{\max} и $|\zeta|_{\max}$ соответственно.

Таблица 3

Максимальные значения повышения (ζ_{\max}) и понижения ($|\zeta|_{\max}$) уровня моря, время их достижения и соответствующие им координаты

Скорость ветра	ζ_{\max} , м	x_{\max} , км	y_{\max} , км	t_{\max} , ч. мин	$ \zeta _{\max}$, м	x_{\max} , км	y_{\max} , км	t_{\max} , ч. мин
$\mathbf{W}_{\text{st}}^1 = 5 \text{ м/с}$	0,20	366,51	219,61	17.10	0,10	4,02	95,07	13.23
$\mathbf{W}_{\text{st}}^2 = 10 \text{ м/с}$	0,69	367,86	218,27	18.46	0,39	4,02	107,13	13.47
$\mathbf{W}_{\text{SKIRON}}$	0,43	278,50	90,10	168.15	0,25	20,10	122,50	190.55
$\mathbf{W}_{\text{st}}^1 + \mathbf{W}_{\text{SKIRO}}\text{N}$	0,55	296,14	100,05	171.11	0,28	22,34	124,14	191.30
$\mathbf{W}_{\text{st}}^2 + \mathbf{W}_{\text{SKIRON}}$	0,72	290,20	101,10	172.20	0,58	21,25	124,00	192.45

Поля уровня моря Азовского моря, полученные в результате моделирования, показаны на рис. 2.

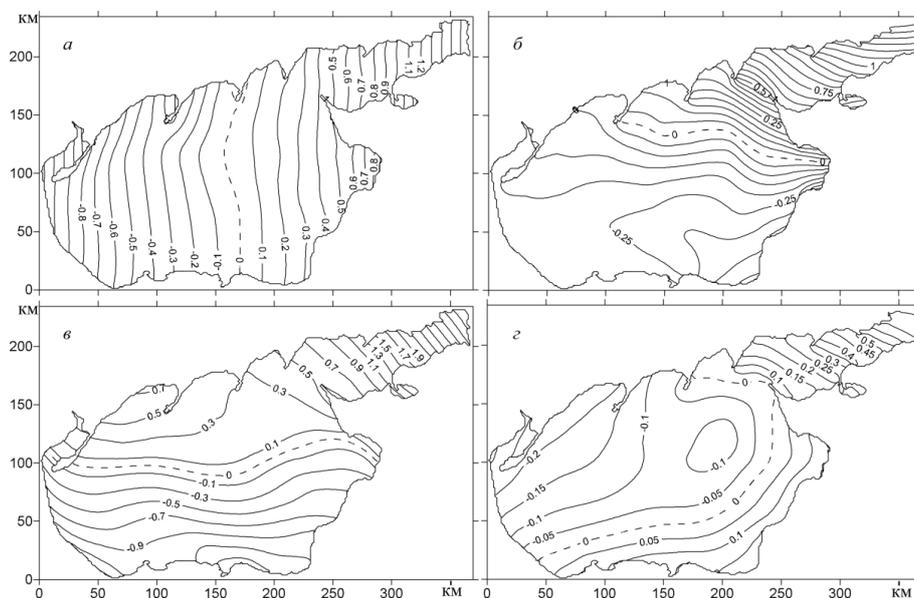


Рис. 2. Поля уровня Азовского моря в различные моменты времени: стационарный режим, 48 ч (а); 68 ч (б); 90 ч (в); 140 ч (г)

В установившемся режиме (рис. 2, а) отмечаются понижения уровня вдоль западного побережья (сгоны) и повышения — вдоль восточного (нагоны). Узловая (штриховая) линия пересекает центральную часть моря, она ориентирована перпендикулярно

направлению ветра. С началом действия нестационарного ветра (рис. 2, б — г) области сгона и нагона перемещаются вдоль центральной части моря.

В табл. 4 для береговых станций Азовского моря приведены расчетные максимумы нагонов и сгонов, вызванных действием ветра $W_{st}^{1,2}$, W_{SKIRON} и $W_{st}^{1,2} + W_{SKIRON}$. Здесь же указано время их достижения. В верхней части находятся величины нагонов, в нижней — сгонов. Из анализа представленных данных следует, что наибольшие нагоны, генерируемые стационарным ветром, имеют место на ст. Таганрог (20,7; 62,4 см), прогностическим ветром — на ст. Приморско-Ахтарск (57,1 см), при суперпозиции ветров — на ст. Приморско-Ахтарск (80,4 см) и Ейск (102,2 см) соответственно. Отсюда видно, что максимум нагона при $W_{st}^2 + W_{SKIRON}$ в 1,27 раза больше, чем при $W_{st}^1 + W_{SKIRON}$. Минимальные нагоны возникают в Мысовом (7,5 см при W_{st}^1 ; 13,9 см при W_{st}^2) и в Опасном (9,4 см при W_{SKIRON} ; 16,1 см при $W_{st}^1 + W_{SKIRON}$ и 34 см при $W_{st}^2 + W_{SKIRON}$).

Таблица 4

Максимальные нагоны и сгоны (см) на береговых станциях Азовского моря при постоянном и прогностическом ветрах, а также при их совместном действии

Береговые станции	W_{st}^1	W_{st}^2	W_{SKIRON} N	$W_{st}^1 + W_{SKIRON}$ N	$W_{st}^2 + W_{SKIRON}$
Геническ	—	—	25,4	32,2	62,3
Бердянск	—	—	9,6	16,9	44,3
Мариуполь	9,8	37,3	29,3	46,4	80,4
Таганрог	20,7	62,4	50,6	63,1	89,5
Ейск	13,8	52,2	38,1	76,0	102,2
П.-Ахтарск	8,1	43,2	57,1	80,4	91,1
Темрюк	10,2	26,9	24,5	29,7	49,9
Опасное	—	—	9,4	16,1	34,0
Мысовое	7,5	13,9	12,1	19,6	34,2
Геническ	12,2	51,7	42,5	76,5	87,0
Бердянск	4,0	17,6	17,3	30,9	62,1
Мариуполь	—	—	18,2	26,0	39,7
Таганрог	—	—	29,0	42,4	72,9
Ейск	—	—	18,9	41,1	45,3
П.-Ахтарск	—	—	14,1	23,8	35,5
Темрюк	—	—	8,7	11,2	33,0
Опасное	3,3	11,1	10,6	20,5	34,2
Мысовое	—	—	22,3	39,4	63,9

В результате совместного воздействия переменного и прогностического ветров максимальные сгоны становятся больше, чем в установившемся режиме и при отсутствии стационарных течений. Для всех рассматриваемых ветров максимальные сгоны имеют место в Геническе: 12,2 см (W_{st}^1), 51,7 см (W_{st}^2), 42,5 см (W_{SKIRON}), 76,5 см ($W_{st}^1 + W_{SKIRON}$) и 87 см ($W_{st}^2 + W_{SKIRON}$). Минимальные сгоны возникают в Опасном: 3,3 см

(W_{st}^1), 11,1 см (W_{st}^2) и в Темрюке: 8,7 см (W_{SKIRON}), 11,2 см ($W_{st}^1 + W_{SKIRON}$) и 33 см ($W_{st}^2 + W_{SKIRON}$).

3.2. Определение размеров областей осушения и затопления побережья Азовского моря в зависимости от действующего ветра

Выполним анализ протяженности территорий возможных осушений и затоплений, возникающих под действием только постоянного ветра, только прогностических полей, и также в результате совместного действия нестационарных возмущений и фоновых стационарных течений. Зная максимальную величину сгона (нагона) и угол наклона (подъема) рельефа береговой зоны (α), определим размер максимально возможной области осушения (затопления), вызванного действием ветра в различных прибрежных районах. Для Азовского моря характерными являются углы опускания и подъема побережья от 1,5 до 3°.

В табл. 5 даны размеры участков затопления (верхняя часть таблицы) и осушения (нижняя часть), вызванных действием стационарного и прогностического ветров в прибрежных районах Азовского моря.

Таблица 5

Размеры областей затопления и осушения (м) в районах станций Азовского моря в зависимости от угла наклона (α) рельефа прибрежной области

Береговые станции	W_{st}^1		W_{st}^2		W_{SKIRON}		$W_{st}^1 + W_{SKIRON}$		$W_{st}^2 + W_{SKIRON}$	
	$a=3^\circ$	$a=1,5^\circ$	$a=3^\circ$	$a=1,5^\circ$	$a=3^\circ$	$a=1,5^\circ$	$a=3^\circ$	$a=1,5^\circ$	$a=3^\circ$	$a=1,5^\circ$
Геническ	—	—	—	—	6,2	12,3	11,9	23,8	4,9	9,7
Бердянск	1,9	3,7	7,1	14,2	3,2	6,5	8,5	16,9	1,7	3,5
Мариуполь	4,0	7,9	11,9	23,8	8,9	17,7	15,4	30,7	5,6	11,2
Таганрог	2,6	5,3	10,0	19,9	12,1	24,1	17,1	34,2	9,7	19,3
Ейск	1,5	3,1	8,3	16,5	14,5	29,0	19,5	39,0	7,3	14,6
П.-Ахтарск	1,9	3,9	5,1	10,3	15,4	30,7	17,4	34,8	10,9	21,8
Темрюк	—	—	—	—	5,7	11,3	9,5	19,1	4,7	9,4
Опасное	—	—	—	—	3,1	6,2	6,5	13,0	1,8	3,6
Мысовое	2,3	4,7	9,9	19,8	3,7	7,5	6,5	13,1	2,3	4,6
Геническ	0,8	1,5	3,4	6,7	14,6	29,2	16,6	33,2	8,1	16,2
Бердянск	—	—	—	—	5,9	11,8	11,9	23,7	3,3	6,6
Мариуполь	—	—	—	—	5,0	9,9	7,6	15,2	3,5	7,0
Таганрог	—	—	—	—	8,1	16,2	13,9	27,8	5,5	11,1
Ейск	—	—	—	—	7,9	15,7	8,7	17,3	3,6	7,2
П.-Ахтарск	—	—	—	—	4,5	9,1	6,8	13,6	2,7	5,4
Темрюк	0,6	1,3	2,1	4,2	2,1	4,3	6,3	12,6	1,7	3,3
Опасное	1,9	3,7	7,1	14,2	3,9	7,8	5,8	11,5	2,0	4,0
Мысовое	—	—	—	—	7,5	15,1	12,2	24,4	4,3	8,5

Из анализа приведенных данных следует, что под действием ветра $\mathbf{W}_{st}^2 + \mathbf{W}_{SKIRON}$ наибольшему затоплению (21,8 м) подвергается район Приморско-Ахтарска. При этом протяженность участка осушения в районе Геническа (16,2 м) превосходит значения, полученные для других прибрежных территорий. Сравнивая величины областей затопления и осушения для различных углов наклона береговой зоны, отметим, что при увеличении этих углов размеры районов затопления и осушения уменьшаются.

Так, если $\alpha = 3^\circ$, наибольшие затопления имеют место при $\mathbf{W}_{st}^{1,2}$ в Мариуполе (4 и 11,9 м), при \mathbf{W}_{SKIRON} — в Приморско-Ахтарске (15,4 м), при $\mathbf{W}_{st}^{1,2} + \mathbf{W}_{SKIRON}$ — в Ейске (19,5 м) и в Приморско-Ахтарске (10,9 м) соответственно. Если $\alpha = 1,5^\circ$, наибольшие затопления возникают при $\mathbf{W}_{st}^{1,2}$ в Мариуполе (7,9 и 23,8 м), при \mathbf{W}_{SKIRON} — в Приморско-Ахтарске (30,7 м), при $\mathbf{W}_{st}^{1,2} + \mathbf{W}_{SKIRON}$ — в Ейске (39 м) и в Приморско-Ахтарске (21,8 м) соответственно. При \mathbf{W}_{SKIRON} и $\mathbf{W}_{st}^{1,2} + \mathbf{W}_{SKIRON}$ наибольшие осушения для всех значений углов α наблюдаются в Геническе. Так, при $\alpha = 1,5^\circ$ размеры этих областей равны 29,2; 33,2 и 16,2 м соответственно.

3.3. Сравнение результатов моделирования отклонений уровня моря, вызываемых действием полей атмосферной модели SKIRON, с данными натурных наблюдений

Сравним результаты моделирования и натурных измерений уровня моря, приведенных в таблицах ежечасных данных высот уровня Государственной метеорологической службой Украины за период 8 — 18 сентября 2007 г. Проведем оценку рассчитанных значений экстремальных отклонений уровня, вызванных только прогностическим полем \mathbf{W}_{SKIRON} и P_{SKIRON} , и ежечасных данных из указанных таблиц. В Геническе величина расчетного максимума составляет 25,4 см, что на 4,7 см (16%) меньше, чем по данным наблюдений. В Мариуполе найденный в результате расчетов максимальный нагон (29,3 см) на 4 см (12%) меньше измеренного. Отсюда следует, что амплитуды колебаний уровня, полученные по натурным данным и путем численных расчетов, отличаются незначительно. При этом разница максимумов сгонов и нагонов может быть объяснена определенными погрешностями при измерениях и вычислениях.

3.4. Исследование влияния фоновых стационарных течений на эволюцию примеси в Азовском море, вызванную действием переменного ветра

Целью следующих численных экспериментов является оценка влияния рассматриваемых полей ветра и возникающих при этом течений на распространение пассивной примеси, поступающей в центральный район моря. Начальное положение выброса на расчетной сетке модели было выбрано при $x_0 = 180$ км, $y_0 = 120$ км, где глубина моря составляет 12 м. Выброс примеси имеет вид цилиндрической области радиусом R и глубиной h_1 ($0 \geq z \geq h_1$, где h_1 — шаг по вертикали в поверхностном слое), заданное соотношением (11). Начальное значение концентрации постоянно в данной области и равно единице ($C(x, y, z, t_0) = C_0(x, y, z) = 1$).

Момент времени выброса загрязнения в экспериментах зависит от полей действующего ветра. Так, для $\mathbf{W}_{st}^{1,2}$ это время совпадает с моментом установления движения жидкости ($t_0 = t_1$), равным 38 и 43 ч соответственно. Для \mathbf{W}_{SKIRON} момент выброса про-

исходит в 0 ч 11 сентября 2007 г. ($t_0 = 72$ ч). Для $\mathbf{W}_{st}^{1,2} + \mathbf{W}_{SKIRON}$ это время определяется следующим образом: $t_0 = t_1 + 72$ ч. При анализе времени распространения и рассеивания примеси принимаем, что во всех случаях момент выброса соответствует начальному значению $t_0 = 0$.

В табл. 6 приведены коэффициент максимального распространения пассивной примеси (K_{max}) с соответствующим ему моментом времени (t_{max} , ч) и время полного рассеивания (t_d , ч) примеси для двух скоростей постоянного ветра, прогностического ветра и их совместного действия. Рассматриваются горизонты ($z = 0$; $z = -H/2$ и $z = -H + h_2$), на которых прослеживаются тенденции эволюции примеси. Согласно приведенным данным, площадь распространения области загрязнения зависит от скорости ветра, приводящего к установившемуся движению. Для большей скорости постоянного ветра (\mathbf{W}_{st}^2) становятся больше скорости течений. (табл. 2), при этом увеличиваются площадь распространения примеси и время ее полного рассеивания (табл. 5). В связи с этим совместное действие стационарного и прогностического ветром также приводит к увеличению области загрязнения.

Из табл. 6 видно, что наибольшая площадь загрязнения имеет место при максимальном значении скорости действующего ветра $\mathbf{W}_{st}^2 + \mathbf{W}_{SKIRON}$. В этом случае на свободной поверхности $K_{max} = 1,32$ через 40ч после выброса примеси. При этом время ее полного рассеивания (t_d) составляет 86,5 ч. Наибольшая площадь области загрязнения на глубине $z = -H/2$ достигается спустя 42 ч после выброса примеси ($K_{max} = 1,35$). Полное рассеивание загрязнения на этой глубине происходит через 110 ч. В придонном слое ($z = -H + h_2$) максимальное значение коэффициента распространения области загрязнения (1,38) отмечено в момент времени $t_{max} = 59$ ч. При этом спустя 115 ч с момента выброса примеси ее концентрация во всей акватории моря не превышает 2,5% от начального значения ($C_d = 2,5 \cdot 10^{-2}$).

Таблица 6

Параметры эволюции распространения примеси на различных глубинах Азовского моря при постоянном ($\mathbf{W}_{st}^1 = 5$ м/с, $\mathbf{W}_{st}^2 = 10$ м/с) и прогностическом ветрах, а также при их совместном действии

Глубина, м	Максимальные значения	\mathbf{W}_{st}^1	\mathbf{W}_{st}^2	\mathbf{W}_{SKIRO} N	$\mathbf{W}_{st}^1 + \mathbf{W}_{SKIRO}$ N	$\mathbf{W}_{st}^2 + \mathbf{W}_{SKIRON}$
$z = 0$	K_{max}	1,14	1,18	1,25	1,30	1,32
	t_{max}	5,7	4,9	31	40	40
	t_d	17,3	18,2	57,1	84,5	86,5
$z = -H/2$	K_{max}	1,16	1,18	1,27	1,33	1,35
	t_{max}	14,3	14,7	34	42	42
	t_d	36,7	37,6	104	106	110
$z = -H + h_2$	K_{max}	1,16	1,19	1,33	1,37	1,38
	t_{max}	26,9	25,5	55	58	59
	t_d	53,4	55,2	108	110	115

Заключение

В работе представлены результаты численных экспериментов расчета сгонно-нагонных явлений и распространения пассивной примеси системой течений, вызванных воздействием постоянного ветра и прогностических полей в Азовском море. Точность полученных результатов подтверждена сравнением значений экстремальных сгонов и нагонов с натурными данными, полученными по измерениям уровня на береговых станциях.

Представлены таблицы величин сгонов и нагонов, скоростей течений, размеров затоплений и осушений, генерируемых действием нестационарного и постоянного ветров. При этом показано, что наибольшие нагоны имеют место в районах Таганрога, Приморско-Ахтарска и Ейска, а максимальные сгоны — в Геническе. Построены карты уровня Азовского моря, сделаны выводы о зависимости эволюции пассивной примеси от скорости стационарного ветра и действия переменного неоднородного ветра. При этом, установлено, что полное рассеивание примеси быстрее происходит в поверхностном слое и медленнее — в придонном. Выполнен анализ протяженности территорий возможных осушений и затоплений в зависимости от рельефа береговой зоны и скорости постоянно-го ветра, а также от его совместного действия с полями атмосферной модели SKIRON.

С учетом того, что в последние годы наблюдается рекордное количество стихийных бедствий, влияющих на безопасность коммуникаций Азово-Черноморского региона, имеется много важных направлений исследований. Прежде всего, это оперативное прогнозирование предвестников экстремальных опасных явлений, происходящих в акватории Азовского моря. Для этого необходимы теоретические исследования, позволяющие определить физические закономерности циркуляции вод Азовского моря на основании анализа экстремальных событий. Большой интерес, также, представляет изучение сейшеобразных колебаний на формирование экстремальных уровней и течений в Азовском море.

Литература

1. Гидрометеорологические условия шельфовой зоны морей СССР. Том III. Азовское море. Л.: Гидрометеиздат, 1986. 218 с.
2. Дацюк В.Н., Крукиер Л.А., Чикин А.Л., Чикина Л.Г. Моделирование экстремального наводнения в дельте Дона на многопроцессорных вычислительных системах // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. 2014. Т. 3, № 1. С. 80–88.
3. Иванов В.А., Черкесов Л.В., Шульга Т.Я. Динамические процессы и их влияние на распространение и трансформацию загрязняющих веществ в ограниченных морских бассейнах. Севастополь: ЭКОСИ-Гидрофизика, 2010. 178 с.
4. Иванов В.А., Фомин В.В., Черкесов Л.В., Шульга Т.Я. Исследование сгонно-нагонных явлений в Азовском море, вызванных атмосферными возмущениями // Доклады НАН Украины. 2006. № 11. С. 109–113.
5. Инжебейкин Ю.И. Особенности формирования кратковременных наводнений и экстремальных течений в Азовском море // Труды Государственного океанографического института. 2011. № 213. С. 91–102.
6. Матишов Г.Г., Бердников С.В., Беспалова Л.А., Ивлиева О.В., Цыганкова А.Е., Хартиев С.М., Иошпа А.Р., Кропянко Л.В., Сушко К. С., Шевердяев И.В., Беспалова Л.А.

- лова Е.В. Современные опасные экзогенные процессы в береговой зоне Азовского моря. Ростов-на-Дону: Издательство Южного федерального университета, 2015. 324 с.
7. Фомин В.В. Численная модель циркуляции вод Азовского моря // Научные труды УкрНИГМИ. 2002. Вып. 249. С. 246–255.
 8. Фомин В.В., Шульга Т.Я. Исследование волн и течений, возникающих под действием ветра в Азовском море // Доклады НАН Украины. 2006. № 12. С. 110–115.
 9. Черкесов Л.В., Иванов В.А., Хартиев С.М. Введение в гидродинамику и теорию волн. СПб.: Гидрометеиздат, 1992. 264 с.
 10. Шабас И.Н., Чикин А.Л., Чикина Л.Г. Математическое моделирование задач переноса многокомпонентных примесей в Азовском море на многопроцессорных вычислительных системах // Известия ЮФУ. Технические науки. 2014. № 12 (161). С. 200–210.
 11. Blumberg A.F., Mellor G.L. A description of three-dimensional coastal ocean-circulation model in Three-Dimensional Coast Ocean Models // Coast. Estuar. Sci. 1987. No. 4. P. 1–16. DOI: 10.1029/co004p0001.
 12. Courant R., Friedrichs K.O., Lewy H. On the partial difference equations of mathematical physics // IBM J. 1967, March. P. 215–234.
 13. URL: <http://forecast.uoa.gr> (дата обращения: 30.08.2016).
 14. Mellor G.L. and Yamada T. Development of a turbulence closure model for geophysical fluid problems // Rev. Geophys. Space Phys. 1982. Vol. 20, No. 4. P. 851–875. DOI: 10.1029/rg020i004p00851.
 15. Smagorinsky J. General circulation experiments with primitive equations, I. The basic experiment // Mon. Wea. Rev. 1963. Vol. 91, No 3. P. 99–164. DOI: 10.1175/1520-0493(1963)091<0099:gcewtp>2.3.co;2.

Черкесов Леонид Васильевич, член-корреспондент НАН Украины, д.ф.-м.н., профессор, главный научный сотрудник, отдел теории волн, Морской гидрофизический институт РАН (Севастополь, Российская Федерация)

Шульга Татьяна Яковлевна, к.ф.-м.н., старший научный сотрудник, отдел теории волн, Морской гидрофизический институт РАН (Севастополь, Российская Федерация)

DOI: 10.14529/cmse170104

STUDY OF STATIONARY CURRENTS FOR DYNAMIC PROCESSES AND ADMIXTURES OF POLLUTION IN THE SEA OF AZOV

© 2017 L. V. Cherkesov, T. Ya. Shul'ga

Federal Federal State Budget Scientific Institution "Marine Hydrophysical Institute of RAS"

(Kapitanskaya, 2, Sevastopol' 299011 Russia)

E-mail: tyshulga@mail.ru

Received: 20.10.2016

With the use of three-dimensional nonlinear mathematical model dynamic processes and features of transformation are studied admixtures in the Sea of Azov, caused the action of variable wind and atmospheric pressure

at presence of stationary currents. On the basis of results of numeral calculations conclusions are done about influence of sizes of speeds of stationary flows on the maximal rejections of level and speed of non-stationary currents, generated the fields of wind, SKIRON got on an atmospheric model.

The analysis of influence of change intensity of stationary flows is executed on the sizes of areas of drainage and submergence in off-shore districts exterminating depending on the angle of slope (getting up) of relief of waterside area. It is set that the united action permanent and wind was simulated with the SKIRON weather forecasting system result in the substantial increase of area of distribution of passive admixture and time of its dispersion as compared to influence only of stationary currents.

Keywords: numerical modeling, sigma-coordinate model, dynamic processes, stationary currents, areas of drainage and submergence, evolution of passive admixture.

FOR CITATION

Cherkesov L.V., Shul'ga T.Ya. Study of stationary currents for dynamic processes and admixtures of pollution in the sea of Azov. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2017. vol. 6, no. 1. pp. 56–72. (in Russian) DOI: 10.14529/cmse170104.

References

1. *Gidrometeorologicheskie usloviya shel'fovoj zony morej SSSR. Tom III, Azovskoe more* [Hydrometeorological conditions of the shelf zone of the seas of the USSR. Volume III, Sea of Azov]. St. Petersburg, Gidrometeoizdat, 1986. 218 p. (in Russian)
2. Datsyuk V.N., Krukier L.A., Chikin A.L., Chikina L.G. Simulation of extreme floods in the delta of the Don on multiprocessor computer systems. *Bulletin of the South Ural State University. Series: Computational Mathematics and Informatics*. 2014. vol. 3, no. 1. pp. 80–88. (in Russian)
3. Ivanov V. A., Cherkesov L. V., Shul'ga T. Ya. Dinamicheskie processy i ih vliyanie na rasprostraneniye i transformaciyu zagryaznyayushchih veshchestv v ogranichennyh morskikh bassejnah [Dynamic processes and their influence on the distribution and transformation of pollutants in restricted marine basins]. Sevastopol, NPC «EKOSI-Gidrofizika», 2010. 178 p. (in Russian)
4. Ivanov V. A., Fomin V. V., Cherkesov L. V., Shul'ga T. Ya. Study of rundown-onset events in the Sea of Azov caused by atmospheric fluctuations. *Dopov. Nats. Akad. Nauk Ukraini*. 2006. no. 11, 109–113.
5. Inzhebeykin Yu.I. Features of formation of short-term flooding and extreme currents in the sea of Azov. *Tr. Gos. Meteorol. Inst.* 2011. vol. 213. pp. 91–102.
6. Matishov G.G., Berdnikov S.V., Bepalova L.A., Ivlieva O.V., Tsyigankova A.E., Hartiev S.M., Ioshpa A.R., Kropyanko L.V., Sushko K, S., Sheverdyayev I.V., Bepalova E.V. Modern hazardous exogenous processes in the coastal zone of Azov Sea. Rostov-on-Don, Publishing of the South Federal University, 2015. 324 p. (in Russian)
7. Fomin V. V. Numerical model of the circulation of the Azov Sea water. *Tr. Ukr. Nauchno-Issled. Gos. Meteorol. Inst.* 2002. vol. 249. pp. 246–255.
8. Fomin V. V., Shul'ga T. Ya. Study of the waves and currents caused by the wind in the Sea of Azov. *Dopov. Nats. Akad. Nauk Ukraini*. 2006. no. 12. pp. 110–115.
9. Shabas I.N., Chikin A.L., Chikina L.G. Mathematical modeling of the problems of multi-component pollution transport in the sea of Azov on the multiprocessor computing systems. *Izvestiya SFedU. Engineering Sciences*. 2014. no. pp. 200–210.

10. Cherkesov L. V., Ivanov V. A., Khartiev S. M. *Vvedenie v gidrodinamiku i teoriiy voln* [Introduction into Hydrodynamics and Wave Theory]. St. Petersburg, Gidrometeoizdat, 1992. 264 p. (in Russian)
11. Blumberg A.F., Mellor G.L. A description of three-dimensional coastal ocean-circulation model in Three-Dimensional Coast Ocean Models. *Coast. Estuar. Sci.* 1987. no. 4. pp. 1–16. DOI: 10.1029/co004p0001.
12. Courant R., Friedrichs K.O., Lewy H. On the partial difference equations of mathematical physics. *IBM J.* 1967, March. pp. 215–234.
13. URL: <http://forecast.uoa.gr> (accessed: 30.08.2016).
14. Mellor G. L., Yamada T. Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.* 1982. vol. 20, no. 3. pp. 851–875. DOI: 10.1029/rg020i004p00851.
15. Smagorinsky J. General circulation experiments with primitive equations, I. The basic experiment. *Mon. Wea. Rev.* 1963. vol. 91, no 3. pp. 99–164. DOI: 10.1175/1520-0493(1963)091<0099:gcewtp>2.3.co;2.

HYBRID COMPUTER SYSTEM PROGRAMMING TECHNOLOGY WITH ADAPTATION AND SCALING OF CALCULATIONS*

© 2017 г. А.А. Gulenok¹, А.И. Dordopulo², И.И. Levin², В.А. Gudkov²

¹ *Academician A.V. Kalyaev SRI multiprocessor computer system
at Southern Federal University (GSP-284, Chekhov st. 2, Taganrog, 347928, Russia),*

² *Scientific Research Centre of Supercomputers and Neurocomputers
(Italyansky lane 106, Taganrog, 347900, Russia),*

*E-mail: andrei_gulenok@mail.ru, scorpio@mvs.sfedu.ru, levin@superevm.ru,
Slava_Gudkov@mail.ru*

Received: 20.10.2016

The paper considers the programming technology for hybrid computer systems, which contain reconfigurable and microprocessor computational nodes. The base of the programming technology for hybrid computer systems is the high-level programming language COLAMO with extensions, which allow descriptions of various types of parallel calculations such as structural, structural-procedural, multi-procedural and procedural forms of organization of calculations in a unified parallel-pipeline form. The suggested parallel-pipeline form allows modifications of forms of organization of calculations. Such modifications are performed automatically by the COLAMO language pre-processor, which takes into account current configuration of the hybrid computer system. Owing to the suggested technology, the program can be automatically adapted to the changed architecture or configuration of the hybrid computer system without any modifications of the source code made by the developer. Specially for this the source parallel program, developed in the programming language COLAMO, is transformed by the pre-processor into the canonical form. Then the pre-processor estimates the available computational resource, detects effective parameters of implementation of the program on the available resource and, if necessary, reduces the program performance to adapt it to the current configuration of the hybrid computer system. The technology provides two-way scaling: for increasing of the available computational resource (induction), and for reducing of the available computational resource (reduction), which provides resource independence of programming during implementation of the program, i.e. the developer is not “bound” to the available hardware resource of the computer system.

Keywords: performance reduction, high-level programming language, programming of hybrid computer systems, application adaptation, application scaling.

FOR CITATION

Gulenok A.A., Dordopulo A.I., Levin I.I., Gudkov V.A. Hybrid computer system programming technology with adaptation and scaling of calculations. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2017. vol. 6, no. 1. pp. 73–86. DOI: 10.14529/cmse170105.

Introduction

The majority of real-world problems require combination of both sequential and parallel computational fragments within a single computational space for effective implementation of structural and procedural [1] fragments of calculations. Many developers consider design of computer systems with hybrid organization of calculations as a solution of this problem. Such computer systems can contain computational nodes with different architectures, united by data

* The paper is recommended for publication by the Program Committee of the International Scientific Conference “Parallel computational technologies (PCT) 2017”.

transfer channels, and allow implementation of structural and procedural calculations within a single computational space. Symbiosis of nodes with different architecture in one computer system theoretically allows the growth of the computer system real performance owing to the opportunity of effective implementation of both structural and procedural fragments of calculations in the nodes with different architecture.

Wide application of such computer systems for solving real-world problems is considerably limited by high complexity of their programming, as the effective use of architectural advantages of all computational nodes requires not only deep knowledge of various programming languages and development environments for designing computational nodes of various types, but also skills of independent synchronization of calculations within a single space.

The paper is organized as follows. In Section 1 we describe the main reasons that have lead us to development of the hybrid computer systems programming technology. Section 2 describes the single parallel-pipeline form of COLAMO-applications developed for HCS, that allows modifications of organization of calculations. In Section 3 we describe performance reduction methods as the base for two-way scaling of calculations in hybrid computer systems. Section 4 describes implementation of the suggested programming technology as mapping of parallel applications on the hardware resource of the hybrid computer system. In conclusion we summarize the main theoretical and experimental results of the developed hybrid computer systems programming technology.

1. Programming of hybrid computer systems

A hybrid computer system (HCS) contains computational nodes with different architecture and organization of calculations. Such hybrid computer systems can contain reconfigurable computational nodes and nodes of general-purpose microprocessors, such as general-purpose processors, graphic processors or accelerators Intel Xeon Phi [2]. Nowadays, in order to program such computer systems (CS) we traditionally use programming technologies of heterogeneous computer systems, such as CUDA [3], OpenACC, OpenCL [4], etc., which are based on extensions of the programming languages C, C++, FORTRAN and which take into account the architecture of special-purpose microprocessor node. These programming technologies have considerable disadvantages such as poor portability of end solutions between CSs with different architectures and configurations, and poor scalability of applications.

The main reason of these disadvantages is the HCS programming approach, which involves task decomposition in separate fragments. Each fragment is implemented in a separate computational node (on a separate device) of the hybrid computer system. So, each occupied CS node is programmed independently, and as a result, each modification of the CS configuration or of the initial application code requires re-decomposition of the task and development of local applications for each node of the CS.

It is possible to formulate principal programming problems of modern HCS which contain reconfigurable and microprocessor computational nodes:

- 1) FPGAs and microprocessors are programmed in different programming languages and independently of each other.
- 2) The application is developed specially for the current HCS configuration, and each modification of the system structure requires modification of the application code.
- 3) The developer is responsible for synchronization of the data flows within the task structure.

- 4) Porting of the application to another system with similar configuration leads to complete re-development of the application.
- 5) The programming and debugging time required for development of the HCS application is about 6–12 months.

That is why HCS programming requires tools for description of various kinds of organization of calculations (a single language for various architectures) and tools for translation of parallel applications, united into the technology of resource independent HCS programming. From our point of view, the technology of resource independent HCS programming is a combination of knowledge, methods, technological approaches and tools, which provides flexible modification and scaling of the application according to a new computational architecture or configuration of the computer system.

In order to provide functioning of general-purpose processor and reconfigurable computational nodes in a single space, we need a new technology of resource independent HCS programming [5], based on the following principles:

- adaptation of the application to the current HCS configuration is performed automatically by a specialized software tool – a pre-processor based on performance reduction methods [5];
- effective parameters of scaling and performance reduction must be determined without any participation of the developer, only by computer-aided programming tools;
- for computer-aided transformation to the current HCS configuration, the application must be represented in a canonical form (a single parallel-pipeline form).

Transformation of the application into the single parallel-pipeline form makes it possible to increase the task parallelism (induction) if hardware resource is growing, and to reduce (reduction) if hardware resource is decreasing, is the base for application of computer-aided tools. To implement the technology of resource independent HCS programming we must choose a programming language, which allows description of various forms of organization of calculations and programming of general-purpose processor and reconfigurable computational nodes in a single computational space.

Specialized high-level languages [6, 7] for reconfigurable computer system (RCS) [8] programming have C-like syntax, which is usual for the majority of PC developers, and differ from each other by semantic features of call and use of operators [1]. To describe parallel processes in RCS, these languages use a C-language paradigm which is initially sequential. Semantic of the C-language is oriented to interaction of sequential processes, and it does not allow the use of all abilities of RCS during development of parallel applications in these languages. This leads to a semantic gap between the initial information graph of the task, its description in the high-level language and its circuit solution generated by the translator. The result of this gap is a considerable decrease in effectiveness of the parallel application – as a rule, the performance is in 3–5 times lower in comparison with applications developed with the use of HDL-languages.

A promising direction for RCS programming is a high-level language COLAMO [1, 5], developed in Scientific Research Institute of Multiprocessor Computer Systems (SRI MCS SFU, Taganrog, Russia). The language COLAMO is used for description of parallel algorithms and generation of special-purpose computing structures within RCS architecture according to the principles of structural-procedural organization of calculations. Each special-purpose computing structure sequentially performs structurally (hardwarily) implemented fragments of the task information graph. Each graph is a computational pipeline of an instruction flow. So, the

RCS application (the RCS task) consists of a structural component, presented as a set of hardwarily implemented fragments of calculations, and of a procedural component – one and the same control program for all structural fragments, which provides sequential change of computing structures and organizes data flows.

In order to implement calculations in general-purpose processors, the language COLAMO contains instructions for description of procedural organization of calculations and provides fast transition from procedural implementation of calculations in general-purpose processors to structural organization of calculations on reconfigurable computational nodes. A structure *Implicit* is used for implicit declaration of organization of calculations (structural or procedural) for the application fragment. Re-declaration of implementation of the structure *Implicit* allows the developer to use procedural organization of calculations instead of structural one, and vice-versa without any considerable modification of the parallel program. Owing to this, the developer can create a single application using one and the same programming language for all HCS nodes. This allows the high-level programming language COLAMO to be considered as a base for the technology of resource independent programming for both reconfigurable computational nodes and general-purpose nodes of the HCS.

However, for effective HCS programming it is necessary to have language tools which allow description of fragments of calculations, which use different frequencies, data delay ratio and digital capacity of processed data. Owing to this, it is possible to scale both fragments and single circuit cores in both cases – when hardware resource is increasing or decreasing, and, in addition, it is also possible to use data with variable capacity for effective use of HCS hardware resource.

2. The single parallel-pipeline form of COLAMO-applications developed for HCS

Here and forth, the parallel-pipeline form for representation of variables and arrays is a description of data, made by means of the high-level programming language COLAMO, which provides both parallel and sequential access at the same time. In the language COLAMO such access types are declared by the keywords Vector (BitVector) and Stream (BitStream). Fig. 1 shows an example of simultaneous use of parallel and sequential data and bit access types to the arrays A, B, and C.

Such a form of applications is a canonical form; it allows automatic modification of the principal parameters of any parallel application such as the number of simultaneously implemented computational subgraphs, the capacity of processed data, the number of operations, etc. Such modification can be performed by the pre-processor tool for adaptation of applications to the HCS without any participation of the developer.

Transformation of the initial application into the canonical form is performed in two steps. In the first step variables and structures of the application are transformed for parallel-pipeline processing on data level, and in the second step – on bit level.

In general, the method of application transformation into the canonical form can be represented as follows:

- 1) All arrays in the initial COLAMO-application are transformed into the parallel-pipeline form. If it is necessary, Vector or Stream access types are added.
- 2) All variables (except for loop counters) of the parallel COLAMO-application are transformed into the format of the Union-structure, which provides both direct access to a

variable according to its type, and parallel (bitvector) and sequential (bitstream) access types.

- 3) All subcadr of the parallel COLAMO-application are transformed into Implicit-structures.
- 4) All structures and operators of the application are transformed according to the modified parameters of the variables.
- 5) For the generated single parallel-pipeline form of the COLAMO-application, a global pre-processor directive of performance reduction of functional blocks, which is equal to 1, is declared.

```

Const N = 10; Const M = 100;
Const Bv = 32; Const Bs = 1;
Type Type32: Integer [BV: BitVector, BS:BitStream] of Int;
Var a,b,c : Array Type32 [N : Vector, M:Stream] Mem;
Var i, j, k, t : Number;
Cadr ExpParallelStream;
  For i := 0 to N-1 do
    For j := 0 to M-1 do
      For k := 0 to Bv-1 do
        For t := 0 to Bs-1 do
          Begin
            c[i,j][k,t] := a[i,j][k,t] - b[i,j][k,t];
          End;
        End;
      End;
    End;
  EndCadr;

```

Fig. 1. Concurrent use of parallel and pipeline types of access to data arrays and digits

Variables are transformed to parallel and sequential access types (a mixed access type) according to the following rules:

- if we use only sequential access to array items, then the parameter Vector, which means parallel access, with the dimension of unity is added to the declaration of the array;
- if we use only parallel access to array items, then the parameter Stream, which means sequential access, with the dimension of unity is added to the declaration of the array;
- if we use mixed access to array items, then no transformations are performed for such array.

So, when the declaration of any array is modified, its dimension is increased. For example, when we declare the array A as:

```
Var A : Array Integer [N : Stream] Mem
```

It is necessary to transform its declaration as follows:

```
Var A : Array Integer [M : Vector, K : Stream] Mem,
```

where $M=1$, $K = N/M$. To provide equivalency of the information graphs of the initial application and the modified one, the initial value of M is 1.

Fig. 2 shows transformation of the initial application a) to the canonical form b). As we can see from the text of the parallel-pipeline application (see Fig. 2b), all linear arrays are transformed into two-dimensional arrays with mixed access type (M : Vector, K : Stream), a new loop operator with a loop counter VC_1 is added, and all references to the variables are modified.

```

Const N = 10;
Var a, b, c, d : Array Integer [N: Vector]Mem;
Var i : Number;
Cadr ExpParallel;
  For i := 0 to N-1 do
    Begin
      If(A[i]>5)
        C[i] :=A[i]-B[i];
      Else
        C[i] :=A[i]+D[i];
    end;
  EndCadr;

```

a) The initial program

```

Const N = 10; Const M = 10; Const K = N/M;
Var a, b, c, d : Array Integer [M: Vector, K: Stream]Mem;
Var i,vc_1 : Number;
Cadr ExpParallelConvData;
  For vc_1 := 0 to K-1 do
    Begin
      For i := 0 to M-1 do
        Begin
          If(A[i,vc_1]>5)
            C[i,vc_1] :=A[i,vc_1] - B[i,vc_1];
          Else
            C[i,vc_1] :=A[i,vc_1] + D[i,vc_1];
          end;
        end;
      end;
    EndCadr;

```

b) The transformed program

Fig. 2. Transformation of the program to the canonical form on data level

Effective adaptation of the application for HCS requires similar transformations for bits.

3. Scaling of calculations in HCS on the base of performance reduction

The base for simple scaling and adaptation of the application in both cases – increasing and decreasing of available hardware resource – is reduction of the application performance [5], which means proportional performance reduction in all the fragments of the task information graph and possible reduction of hardware resources necessary for implementation of the computing structure. Performance reduction of parallel applications allows variation of their key parameters (the number of used circuit cores, the number of memory channels, the operand capacity, the frequency, etc.), because structural implementation of the task can lead to a lack of available hardware resource. This is especially urgent if the task is ported between HCS with different architectures and configurations. In contrast to traditional technologies and methods of multiprocessor computer system programming (MPI, CUDA, OpenACC, etc.), in which the basic information subgraph of the task is parallelized depending on the configuration of available hardware resource, the HCS programming technology, in order to apply performance reduction, requires description of the information graph of the task represented in the initial parallel form with the maximum possible degree of parallelism. According to the number of

available HCS computational nodes, the initial information graph is reduced by special reduction transformations, which in a balanced manner reduce performance of all fragments of the information graph and, in several cases, reduce HCS hardware resource used by the task. It is possible to single out four kinds of reduction:

- performance reduction according to circuit operations;
- performance reduction according to memory channels;
- performance reduction according to capacity;
- performance reduction according to frequency.

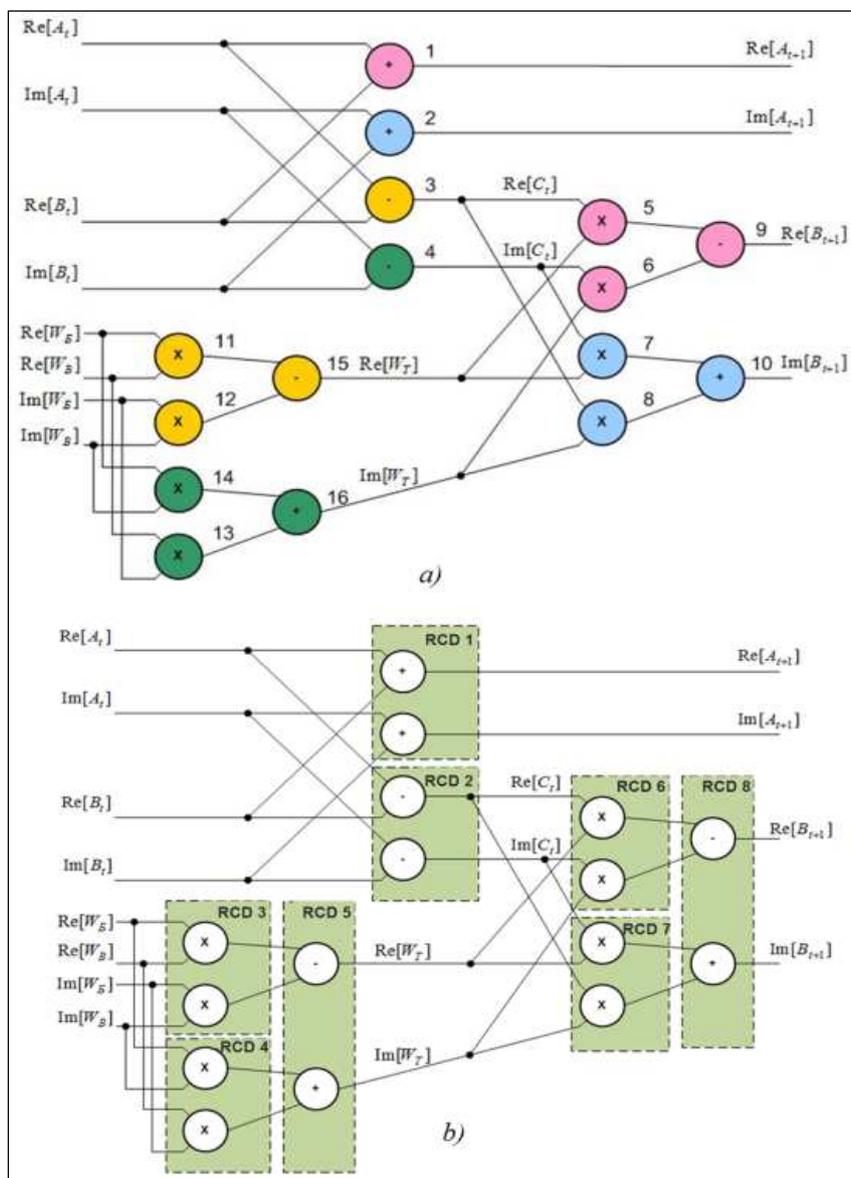


Fig. 3. Principles of performance reduction according to circuit operations illustrated by the operation of fast Fourier transformation (a – the initial information graph of the FFT operation, b – structural implementation of the FFT operation on RCS with conditional filling of FPGAs)

Performance reduction according to circuit operations is based on reducing the number of circuit blocks operating simultaneously and performing computing operations. The example of

the fast Fourier transformation (FFT), shown in Fig. 3 and Fig. 4, illustrates principles of performance reduction according to circuit operations.

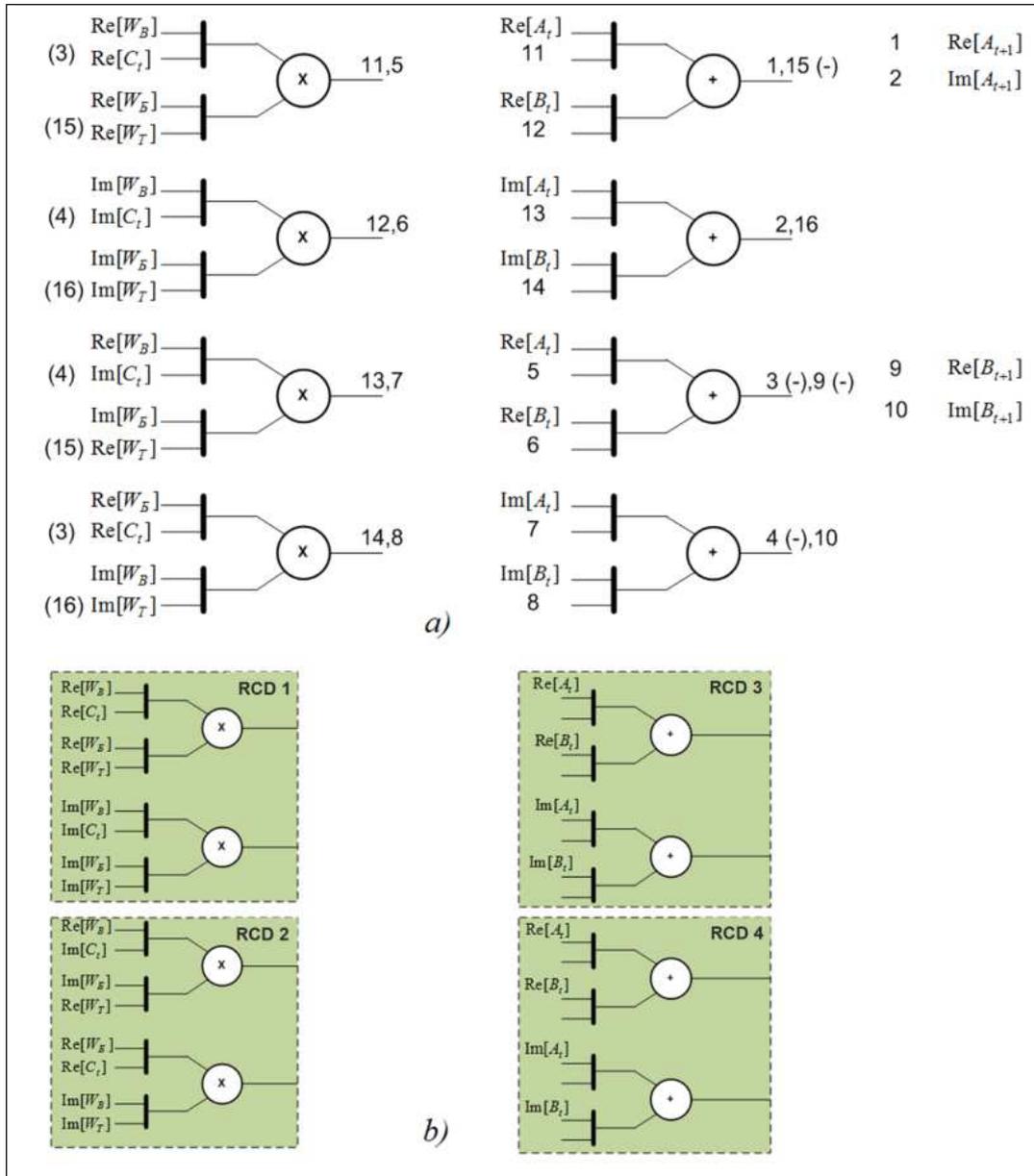


Fig. 4. Result of performance reduction according to circuit operations with the degree equal to 2 for the basic operation of the FFT (a – the information graph of the FFT reduced in 2 times, b – structural implementation of the FFT on RCS with conditional filling of FPGAs)

One of the most important types of reduction, which provide HCS resource independent programming, is performance reduction according to memory channels – an operation of concerted reduction of the number of concurrently used memory channels for some fragment of the task information graph. Fig. 5 and Fig. 6 illustrate principles of performance reduction according to memory channels.

Fig. 5 shows a fragment of the initial information graph, which contains 4 input channels. Fig. 6 shows the result of performance reduction according to memory channels with the degree

equal to 2. After performance reduction according to memory channels the number of concurrently performed circuit operations is not changed but the time of processing data flows increases in 2 times. This decreases the performance of the fragment in 2 times.

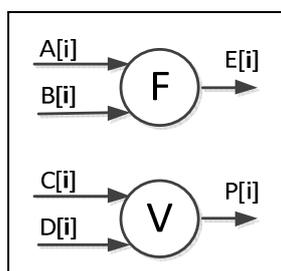


Fig. 5. An initial information graph of some reducible fragment

The result of performance reduction according to capacity is not the reduced number of operations in a computing structure, but the reduced capacity of processing data owing to the use of operations with smaller capacity. This leads to increasing of the processing time and to decreasing of hardware burden for implementation of the computing structure. After performance reduction according to capacity, data flows are controlled by a multiplexer, and data flows enter the computing structure with the data delay ratio equal to the degree of performed reduction.

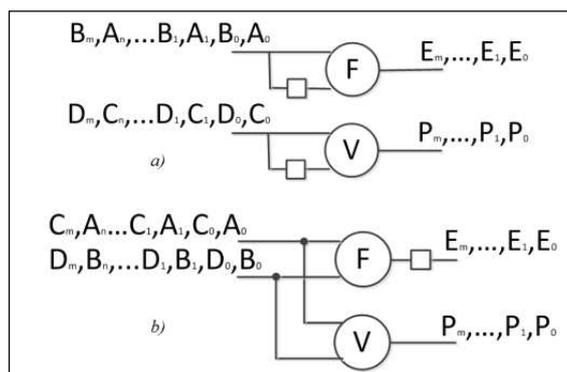


Fig. 6. Result of performance reduction according to memory channels with the degree equal to 2 (a – performance reduction according to memory channels with data placement in one channel for each circuit operation, b – performance reduction according to memory channels with data placement in different channels for each circuit operation).

Performance reduction according to frequency is applied for increasing of the data flow processing time proportionally to the degree of reduction at the cost of multiple reduction of the fragment frequency. Here, the delay ratio of data, entering the computing structure, remains constant. Performance reduction is an auxiliary reduction transformation, which cannot be used separately, and is only applied for matching processing rates between reduced and non-reduced fragments of the information graph in the task structure.

For practical use of the considered reduction transformations the developer must mark the fragments of the parallel application, written in the high-level programming language COLAMO, which can be reduced, by pre-processor directives. The directive of reduction can be declared as follows:

```
#Reduction of <type of reduction> <degree of reduction>;
    Block of operators
EndReduction;
```

During translation of the parallel application the pre-processor automatically transforms the information graph of the application to the available HCS configuration according to the reduction directives placed by the developer. This provides adaptation of the application to the current HCS configuration without considerable modification of the initial program code.

In conclusion of analysis of the transformations, we can describe functioning of the technology of resource independent programming of hybrid computer systems, which contain reconfigurable and microprocessor computational nodes, as follows. The pre-processor unit of the language COLAMO, which performs analysis of the source parallel application, transforms it into the canonical form, determines hardware resource required for its implementation and compares it with the current HCS configuration. Then it determines the maximum reduction degree and types of all required transformations. If the current HCS hardware resource is sufficient for implementation of the application, then bitstream files and loadable files are generated for all HCS nodes involved into implementation of the task. Otherwise special reduction transformations are performed. They reduce the application performance and the involved HCS hardware resource in a balanced manner. Performance reduction is performed as follows: reduction according to simultaneously performed subgraphs of the application, reduction according to capacity, reduction according to instructions (cores), reduction according to data delay ratio (clock rate). The analysis unit of the pre-processor reduces the involved hardware resource for each type of reduction, taking into account its theoretically permissible degree. It determines the most reasonable use of reduction which can provide the maximum possible performance of the application for the current HCS configuration. The text of the reduced parallel application, generated by the pre-processor in an automatic mode, is passed to the translator of the programming language COLAMO, which generates a detailed information graph of the application. The information graph of the application, which contains fragments structurally implemented in reconfigurable computational nodes and procedurally implemented in microprocessor computational nodes, is passed to a synthesizer tool, which automatically distributes the fragments among reconfigurable and microprocessor computational nodes available in the current HCS configuration.

4. Mapping of parallel applications on HCS hardware resource

The problem of distribution of parallel application fragments on hardware resource of a multiprocessor HCS is in automatic decomposition of computational structure of the parallel application, described by the information graph, into disjoint fragments, and in distribution of these fragments into separate computational nodes of the HSC, such as reconfigurable computational nodes and microprocessor computational nodes). Besides, the fragments placed in reconfigurable computational nodes are decomposed into smaller fragments, each of which is implemented in a separate FPGA chip.

All blocks of the computational structure of the parallel application can be divided into two groups. The first group contains hardware implemented blocks, which must be mapped into reconfigurable computational nodes. The second group contains blocks, which correspond to the *Implicit* structures. Each block has its own C++ or C# procedure. The blocks from the second group are placed into HCS microprocessor computational nodes.

Automatic mapping of the computational structure of the parallel application on HCS hardware resource consists of three steps. The first step is decomposition of the computational

structure of the parallel application into disjoint fragments, which must be placed in the nodes of the multiprocessor HCS. Only one block is placed in each microprocessor node of the computational structure, because in one node it is possible to run only one sequential subroutine with intensive data exchange with other microprocessor and reconfigurable nodes. Fragments of parallel applications, which contain hardware implemented blocks, are placed into reconfigurable computational nodes. In this case, the total hardware resource occupied by the blocks of one fragment must not exceed the total hardware resource of FPGA chips of one reconfigurable computational node.

The second step of the algorithm of application synthesis is distribution of the fragments, which are placed in reconfigurable computational nodes, into separate FPGA chips of these nodes.

The third and the last step of the algorithm of application synthesis is synchronization of external and internal data flows of the computational structure of the parallel application, and placement of interface units for matching data exchange between heterogeneous fragments of the computational structure. These interfaces match data exchange between procedural and pipeline computational nodes, between the nodes which are operating at different clock frequencies, between the nodes which are operating at similar clock frequencies but connected to different clock generators (because of different phases and inaccuracy of clock generators), etc.

After setting of all required interfaces and synchronization elements, the synthesizer tool generates bitstream files *.bit for reconfigurable nodes and loadable files *.exe for microprocessor nodes of the HCS, and generates a control program which controls computational process and is single for all HCS computational modules.

Conclusion

For effective programming of hybrid computer systems we suggest the high-level programming language COLAMO as a part of the developed technology of resource-independent programming. The language COLAMO allows description of various forms of organization of calculations in one and the same computational space. The suggested single parallel-pipeline form of applications along with the developed performance reduction methods provide automatic adaptation of applications to the modified HCS architecture or configuration. Owing to the suggested technology we can reasonably use resources of nodes with different architectures during HCS programming, and we have a set of necessary tools for quick development of effective resource-independent scalable parallel applications in a single language space. It simplifies HCS programming and speeds up development of parallel applications.

We tested the developed HCS programming technology, using three test tasks from different problem areas: digital signal processing, symbolic processing and monitoring of computer networks.

The task of digital signal processing consists in processing of input data according to the direct fast Fourier transformation algorithm with frequency decimation. The task of symbolic processing consists in transformation of input data according to the algorithm of symmetric block encryption GOST 28147-89. In the third task we implement a procedure of patterns search in a data flow, which detects frequencies of all found patterns within one string and within the input data flow.

For all three tasks we developed applications in the language COLAMO, and each application was translated for different HCS architectures with different quantity of computational

nodes. Our research has proved effectiveness of the developed technology for adaptation of the parallel application to the used hardware resource in both cases – when resource is increasing and when the application is ported to the HCS with smaller hardware resource.

The project has been funded in part by the scholarship of the President of the Russian Federation for young scientists and graduate students (SP-173.2016.5).

References

1. Kalyaev I.A., Levin I.I., Semernikov E.A., Shmoilov V.I. Reconfigurable multipipeline computing structures. New York, Nova Science Publishers, 2012. 330 p.
2. Dong X., Chai J., Yang J., Wen M., Wu N., Cai X., Zhang C., Chen Z. Utilizing multiple xeon Phi coprocessors on one compute node. *Lecture Notes in Computer Science*, 2014, Vol. 8631, Issue PART 2, pp. 68–81. DOI:10.1007/978-3-319-11194-0_6
3. Liang T.-Y., Li H.-F., Lin Y.-J., Chen B.-S. A Distributed PTX Virtual Machine on Hybrid CPU/GPU Clusters. *Journal of Systems Architecture*, 2016. Vol. 62. pp. 63–77. DOI: 10.1016/j.sysarc.2015.10.003
4. Li H.-F., Liang T.-Y., Lin Y.-J. An OpenMP programming toolkit for hybrid CPU/GPU clusters based on software unified memory. *Journal of Information Science and Engineering*, 2016, Vol. 32, Issue 3. pp. 517–539.
5. Dordopulo A., Levin I., Kalyaev I., Gudkov V., Gulenok A.. Programming of hybrid computer systems based on the performance reduction method. *CEUR Proceedings*, 2016, Vol. 1576, pp. 131–140.
6. El-Araby E., Taher M., Abouellail M., El-Ghazawi T., Newby G.B. Comparative analysis of high level programming for reconfigurable computers: Methodology and empirical study. *2007 3rd Southern Conference on Programmable Logic*, Mar del Plata; 2007; pp. 99–106. DOI: 10.1109/SPL.2007.371731
7. Xu J, Subramanian N, Alessio A, Hauck S. Impulse C vs. VHDL for accelerating tomographic reconstruction. *18th IEEE International Symposium on Field-Programmable Custom Computing Machines*, 2010, pp. 171–174. DOI: 10.1109/fccm.2010.33
8. Dordopulo A., Kalyaev I., Levin I., Slasten L. High-performance reconfigurable computer systems. *Lecture Notes in Computer Science*, 2011, Vol. 6873. Chapter Parallel Computing Technologies. pp. 272–283. DOI:10.1007/978-3-642-23178-0_24

Gulenok Andrey Aleksandrovich, PhD, Senior staff scientist, Academician A.V. Kalyaev SRI multiprocessor computer system at Southern Federal University (Taganrog, Russian Federation)

Dordopulo Alexey Igorevich, PhD, Head of Department, Scientific Research Center of Supercomputers and Neurocomputers (Taganrog, Russian Federation)

Levin Ilya Izrailevich, Dr. Sc., professor, Director of Scientific Research Center of Supercomputers and Neurocomputers (Taganrog, Russian Federation)

Gudkov Vyacheslav Aleksandrovich, PhD, Senior staff scientist, Scientific Research Center of Supercomputers and Neurocomputers (Taganrog, Russian Federation)

ТЕХНОЛОГИЯ ПРОГРАММИРОВАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ ГИБРИДНОГО ТИПА С АДАПТАЦИЕЙ И МАСШТАБИРОВАНИЕМ ВЫЧИСЛЕНИЙ

А.А. Гуленок¹, А.И. Дордопуло², И.И. Левин², В.А. Гудков²

¹ Научно-исследовательский институт многопроцессорных вычислительных систем Южного федерального университета (347928 Таганрог, ул. Чехова, д. 2, ГСП-284)

² Научно-исследовательский центр супер-ЭВМ и нейрокомпьютеров (347900 Таганрог, Итальянский пр., д. 106)

Поступила в редакцию: 20.10.2016

В статье рассматривается технология программирования вычислительных систем гибридного типа, содержащих реконфигурируемые и микропроцессорные вычислительные узлы. В качестве основы технологии программирования вычислительных систем гибридного типа предлагается язык программирования высокого уровня COLAMO с расширениями, с помощью которых можно описывать различные виды параллельных вычислений – структурную, структурно-процедурную, мультипроцедурную и процедурную формы организации вычислений в единой параллельно-конвейерной (канонической) форме. Предложенная параллельно-конвейерная форма позволяет изменять формы организации вычислений автоматизировано препроцессором языка COLAMO с учетом текущей конфигурации вычислительной системы гибридного типа. На основе канонической формы и возможностей описания различных форм организации вычислений на языке программирования высокого уровня COLAMO предложена технология ресурсонезависимого программирования, которая позволяет адаптировать программу под изменившуюся архитектуру или конфигурацию вычислительной системы гибридного типа в автоматическом режиме без корректировки кода программистом. Для этого исходная параллельная программа на языке программирования COLAMO препроцессором преобразуется в каноническую форму, после чего препроцессор проводит оценку доступного вычислительного ресурса, определяет эффективные параметры реализации программы на доступном ресурсе и, при необходимости, выполняет редукцию производительности программы для адаптации под текущую конфигурацию вычислительной системы гибридного типа. Технология позволяет осуществлять масштабирование в обе стороны как в случае увеличения доступного вычислительного ресурса (индукция), так и в случае сокращения доступного вычислительного ресурса (редукция), что обеспечивает ресурсонезависимость программирования при разработке программы — программист не привязывается к доступному аппаратному ресурсу вычислительной системы.

Ключевые слова: редукция производительности, язык программирования высокого уровня, программирование вычислительных систем гибридного типа, адаптация программы, масштабирование программы.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Gulenok A.A., Dordopulo A.I., Levin I.I., Gudkov V.A. Hybrid computer system programming technology with adaptation and scaling of calculations // Вестник ЮУрГУ Серия: Вычислительная математика и информатика. 2017. Т. 6. № 1. С. 73–86. DOI: 10.14529/cmse170105.

Литература

1. Kalyaev I.A., Levin I.I., Semernikov E.A., Shmoilov V.I. Reconfigurable multipipeline computing structures. New York, Nova Science Publishers, 2012. 330 p.
2. Dong X., Chai J., Yang J., Wen M., Wu N., Cai X., Zhang C., Chen Z. Utilizing multiple Xeon Phi coprocessors on one compute node // *Lecture Notes in Computer Science*, 2014, Vol. 8631, Issue PART 2, P. 68–81. DOI:10.1007/978-3-319-11194-0_6
3. Liang T.-Y., Li H.-F., Lin Y.-J., Chen B.-S. A Distributed PTX Virtual Machine on Hybrid CPU/GPU Clusters // *Journal of Systems Architecture*, 2016. Vol. 62. P. 63–77. DOI: 10.1016/j.sysarc.2015.10.003
4. Li H.-F., Liang T.-Y., Lin Y.-J. An OpenMP programming toolkit for hybrid CPU/GPU clusters based on software unified memory // *Journal of Information Science and Engineering*, 2016, Vol. 32, Issue 3. P. 517–539.
5. Dordopulo A., Levin I., Kalyaev I., Gudkov V., Gulenok A.. Programming of hybrid computer systems based on the performance reduction method // *CEUR Proceedings*, 2016, Vol. 1576. P. 131–140.
6. El-Araby E., Taher M., Abouellail M., El-Ghazawi T., Newby G.B. Comparative analysis of high level programming for reconfigurable computers: Methodology and empirical study // *2007 3rd Southern Conference on Programmable Logic, Mar del Plata; 2007*; P. 99–106. DOI: 10.1109/SPL.2007.371731
7. Xu J, Subramanian N, Alessio A, Hauck S. Impulse C vs. VHDL for accelerating tomographic reconstruction // *18th IEEE International Symposium on Field-Programmable Custom Computing Machines*, 2010, P. 171–174. DOI: 10.1109/fccm.2010.33
8. Dordopulo A., Kalyaev I., Levin I., Slasten L. High-performance reconfigurable computer systems // *Lecture Notes in Computer Science*, 2011, Vol. 6873. Chapter Parallel Computing Technologies. P. 272–283. DOI:10.1007/978-3-642-23178-0_24.

ПАРАЛЛЕЛЬНАЯ РЕАЛИЗАЦИЯ СТОХАСТИЧЕСКОЙ КЛЕТОЧНО-АВТОМАТНОЙ МОДЕЛИ РЕКОМБИНАЦИИ ЭЛЕКТРОНОВ И ДЫРОК В 2D И 3D НЕОДНОРОДНЫХ ПОЛУПРОВОДНИКАХ*

© 2017 г. К.К. Сабельфельд, А.Е. Киреева

Институт вычислительной математики и математической геофизики

Сибирского отделения Российской академии наук

(630090 Новосибирск, пр. Академика Лаврентьева, д. 6)

E-mail: karl@osmf.ssc.ru, kireeva@ssd.ssc.ru

Поступила в редакцию: 28.10.2016

В работе представлены стохастические клеточно-автоматные модели рекомбинации электронов и дырок в неоднородном полупроводнике для двумерного и трехмерного случаев. С помощью разработанных клеточно-автоматных моделей рекомбинации исследовано пространственно-временное распределение частиц, обнаружено и исследовано формирование макрокластеров электронов и дырок. В связи с тем, что интегральные характеристики процесса рекомбинации вычисляются с помощью осреднения по большому ансамблю начальных данных, для сокращения времени вычислений разработаны параллельные программы, реализующие клеточно-автоматные модели рекомбинации в двумерном и трехмерном случаях. Параллельная реализация программ позволила вычислить за приемлемое время интегральные характеристики процесса: плотности частиц и интенсивность фотолюминесценции, для большого числа различных начальных условий, а также изучить кинетику процесса рекомбинации при наличии центров рекомбинации и диффузии частиц в двумерном и трехмерном случаях.

Ключевые слова: рекомбинация электронов и дырок, полупроводник, фотолюминесценция, стохастический клеточный автомат, параллельная реализация.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Сабельфельд К.К., Киреева А.Е. Параллельная реализация стохастической клеточно-автоматной модели рекомбинации электронов и дырок в 2D и 3D неоднородных полупроводниках // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. Т. 6, № 1. С. 87–103. DOI: 10.14529/cmse170106.

Введение

Широкозонный полупроводник нитрид галлия (GaN) и его твердые растворы являются одним из самых востребованных и перспективных материалов в области разработок полевых транзисторов и современной оптоэлектроники [1]. Физические свойства GaN и его соединений обеспечивают создание на их основе приборов с улучшенными значениями мощности, напряжения и тока. Для увеличения надежности и эффективности полупроводниковых приборов на основе GaN , а также для расширения области их применения, необходимо исследовать процессы, протекающие в полупроводниках, и изучать характеристики этих процессов в зависимости от параметров GaN .

Изучению зависимости интенсивности фотолюминесценции от структуры полупроводника и внешних параметров посвящены работы российских и зарубежных исследовательских групп. Фотолюминесценция является одной из основных характеристик, измеряемых

*Статья рекомендована к публикации программным комитетом Международной научной конференции «Параллельные вычислительные технологии (ПаВТ) 2016».

при исследовании свойств полупроводниковых приборов. В [2] влияние температуры и плотности оптического возбуждения на мощность лазеров на основе нитрида галлия анализируется с помощью измерения спектров фотолюминесценции при различных параметрах. В работе [3] немецкой группы из Института твердотельной электроники им. П. Друде исследуется рост нановискеров *GaN* путем выявления зависимости интенсивности фотолюминесценции от диаметра и длины отдельного нановискера. В [4] делаются выводы об оптических свойствах выращенного кристалла *GaN* на основе анализа спектра фотолюминесценции при различной концентрации примеси кремния.

Для описания динамики процессов, протекающих в полупроводниках, существуют различные физико-математические модели, основанные на дифференциальных уравнениях и стохастических методах. Работа [5] посвящена моделированию переноса зарядов и рекомбинации электронов и дырок в светодиодах, использующих нитриды третьей группы, к которым относится *GaN*. Одномерная модель переноса зарядов основана на уравнениях Больцмана, Пуассона и Шредингера, которые решаются методом конечных элементов. Полученные результаты сопоставляются с экспериментальными данными. В [6] разработана стохастическая модель транспорта электронов в полупроводнике с учетом его зонной структуры. В этой работе с помощью метода Монте-Карло моделируется движение зарядов с учетом взаимного влияния друг на друга электронов и фононов, а также воздействия ионизации на энергию электронов. Дополнительно, для моделирования электрического поля с помощью разностной схемы решается уравнение Пуассона. Построенная модель позволила вычислить скорости движения зарядов, среднюю энергию электронов, плотность, а также пространственные распределения частиц для различных параметров полупроводников в зависимости от температуры и электрического поля.

Рекомбинацию электронов и дырок можно рассматривать как бимолекулярный процесс [7]. Изучению бимолекулярных процессов посвящены работы нескольких исследовательских групп, в том числе работы американского ученого Д.Т. Джиллеспы. В [8] рассматривается стохастический подход к моделированию реакционно-диффузионного взаимодействия молекул двух видов в хорошо перемешанном разреженном газе, и выводится формула для вычисления вероятности взаимодействия молекул двух разных типов.

Стохастическому моделированию реакционно-диффузионных бимолекулярных процессов посвящены работы К.К. Сабельфельда. В отличие от упомянутых выше моделей, где рассматриваются преимущественно однородные по пространству процессы, в работах К.К. Сабельфельда развивается общий подход, который применим к неоднородным по пространству задачам. В [9–12] разработана стохастическая модель взаимодействия электронов и дырок, основанная на стохастических интегро-дифференциальных уравнениях типа Смолуховского с учетом пространственной неоднородности, диффузии, наличия центров рекомбинации и возможности флуктуаций начальных концентраций. В этих работах изучается процесс рекомбинации электронов и дырок в неоднородном двумерном полупроводнике *GaN* для следующего случая: с помощью фемтосекундного лазера в полупроводнике создается избыток электронов и дырок, которые аннигилируют с друг другом путем туннелирования, процесс ускоряется диффузией и одновременно замедляется нерадиационной рекомбинацией электронов и дырок в дефектах (рекомбинационных центрах). На основе [9, 12] разработана клеточно-автоматная (КА) модель рекомбинации электронов и дырок в неоднородном полупроводнике для двумерного и трехмерного случаев.

Компьютерное моделирование процессов в полупроводнике требует использования значительных вычислительных мощностей, так как необходимо моделировать поведение большого количества частиц в объеме большого размера в течение длительного времени. Кроме того, интегральные характеристики процесса рекомбинации: плотности частиц и интенсивность фотолюминесценции, вычисляются с помощью осреднения по большому ансамблю начальных данных. Для сокращения времени КА-моделирования и вычисления интегральных характеристик процесса разработаны параллельные программы, реализующие КА-модель рекомбинации в двумерном и трехмерном пространстве. С помощью разработанных программ изучается кинетика процесса аннигиляции электронов и дырок при наличии центров рекомбинации и диффузии частиц в двумерном и трехмерном случаях.

Статья включает в себя четыре раздела. В разделе 1 описана математическая постановка задачи, приведена система уравнений Смолуховского и формулы, описывающие асимптотическое поведение характеристик процесса рекомбинации при больших временах. Раздел 2 посвящен описанию КА-модели рекомбинации, в нем объясняются основные термины теории клеточных автоматов. В разделе 3 приводятся результаты распараллеливания КА-модели рекомбинации для трехмерного и двумерного случаев. В разделе 4 представлены результаты КА-моделирования рекомбинации, приведены эволюции КА и значения характеристик для различных параметров моделирования. В заключении статьи перечислены основные результаты, полученные в данной работе, и сделаны выводы о влиянии диффузии, туннелирования и центров рекомбинации на кинетику процесса рекомбинации, промоделированного для двумерного и трехмерного случаев.

1. Математическая модель рекомбинации электронов и дырок в полупроводнике

Для описания процесса рекомбинации электронов и дырок в полупроводнике в [9, 12] предложена стохастическая модель, основанная на системе уравнений Смолуховского (1):

$$\begin{aligned}
 \frac{\partial \rho_n(\mathbf{r}; t)}{\partial t} &= D_n \Delta \rho_n(\mathbf{r}; t) - \rho_n(\mathbf{r}; t) \int B(|\mathbf{x}|) \rho_p(\mathbf{r} + \mathbf{x}; t) d\mathbf{x} \\
 &\quad - \beta^d \rho_n(\mathbf{r}; t) \rho_p(\mathbf{r}; t) - \rho_n(\mathbf{r}; t) \int b_n(|\mathbf{x}|) \rho_{N_n}(\mathbf{r} + \mathbf{x}; t) d\mathbf{x}; \\
 \frac{\partial \rho_p(\mathbf{r}; t)}{\partial t} &= D_p \Delta \rho_p(\mathbf{r}; t) - \rho_p(\mathbf{r}; t) \int B(|\mathbf{x}|) \rho_n(\mathbf{r} + \mathbf{x}; t) d\mathbf{x} \\
 &\quad - \beta^d \rho_p(\mathbf{r}; t) \rho_n(\mathbf{r}; t) - \rho_p(\mathbf{r}; t) \int b_p(|\mathbf{x}|) \rho_{N_p}(\mathbf{r} + \mathbf{x}; t) d\mathbf{x}; \\
 \frac{\partial \rho_{N_n}(\mathbf{r}; t)}{\partial t} &= -\rho_n(\mathbf{r}; t) \int b_n(|\mathbf{x}|) \rho_{N_n}(\mathbf{r} + \mathbf{x}; t) d\mathbf{x} + \rho_p(\mathbf{r}; t) \int b_p(|\mathbf{x}|) \rho_{N_p}(\mathbf{r} + \mathbf{x}; t) d\mathbf{x}.
 \end{aligned} \tag{1}$$

Предполагается, что в начальный момент времени электроны, дырки и рекомбинационные центры случайно распределены в d -мерном пространстве X с плотностями $\rho_n(\mathbf{r}; t)$, $\rho_p(\mathbf{r}; t)$ и $\rho_N(\mathbf{r})$, соответственно. Символ r обозначает пространственную координату, а t — момент времени. Рекомбинационные центры могут быть свободны, тогда они способны захватить электрон, либо заняты электроном, тогда они способны захватить дырку. Суммарная плотность рекомбинационных центров $\rho_N(\mathbf{r}) = \rho_{N_n}(\mathbf{r}; t) + \rho_{N_p}(\mathbf{r}; t)$ в процессе моделирования остается постоянной, меняются только плотности рекомбинационных центров, свободных для электронов $\rho_{N_n}(\mathbf{r}; t)$, и центров, свободных для дырок $\rho_{N_p}(\mathbf{r}; t)$.

Радиационная рекомбинация электронов и дырок осуществляется путем туннелирования со скоростью $B(|\mathbf{x}|) = B_0 \cdot \exp(-|\mathbf{x}|/a_{np})$, где $|\mathbf{x}|$ — расстояние между электроном и дыркой, a_{np} — характерное расстояние взаимодействия электронов и дырок. При радиационной рекомбинации происходит выделение энергии в виде фотона. Захват электронов в свободные рекомбинационные центры происходит со скоростью $b_n(|\mathbf{x}|) = b_{n0} \cdot \exp(-|\mathbf{x}|/a_{nN_n})$. Аналогично, скорость захвата дырок в рекомбинационные центры, содержащие электроны, равна $b_p(|\mathbf{x}|) = b_{p0} \cdot \exp(-|\mathbf{x}|/a_{pN_p})$. При взаимодействии электронов и дырок в рекомбинационных центрах происходит нерадиационная рекомбинация с выделением тепловой энергии.

Электроны и дырки могут диффундировать в области с коэффициентами диффузии D_n и D_p . При перемещении вследствие диффузии встретившиеся электроны и дырки рекомбинируют с друг другом с коэффициентом $\beta^3 = 4\pi \cdot D \cdot r_0$ в трехмерном пространстве и $\beta^2 = 4\pi \cdot D$ в двумерном пространстве, где $D = D_n + D_p$ — это относительный коэффициент диффузии, r_0 — радиус частицы, далее $r_0 = 1$.

Поток фотонов $\phi(t)$ согласно [12] вычисляется по формуле:

$$\phi(t) = \left\langle \int \frac{1}{|X|} d\mathbf{r} \int B(|\mathbf{x}|) \rho_n(\mathbf{r}; t) \rho_p(\mathbf{r} + \mathbf{x}; t) d\mathbf{x} \right\rangle, \quad (2)$$

где угловые скобки обозначают математическое ожидание по начальным распределениям частиц, $|X|$ — размер области. Интенсивность фотолюминесценции $I(t)$ пропорциональна потоку фотонов $\phi(t)$: $I(t) = \hbar\omega \cdot \phi(t)$, где $\hbar\omega$ — энергия фотона [12].

В [12] с помощью корреляционного анализа выведены формулы для асимптотического поведения плотности электронов и интенсивности фотолюминесценции при $t \rightarrow \infty$:

1. Для режима чистого туннелирования, то есть радиационной рекомбинации без рекомбинационных центров и диффузии частиц:

$$\rho'_{n_dD}(t) \sim 1/(\ln(t))^{d/2}; \quad I'_{dD}(t) \sim 1/\{t \cdot (\ln(t))^{d/2+1}\}. \quad (3)$$

2. Для режима чистой диффузии, то есть радиационной рекомбинации вследствие только диффузии частиц:

$$\rho'_{n_dD}(t) \sim 1/t^{d/4}; \quad I'_{dD}(t) \sim 1/t^{d/4+1}, \quad (4)$$

где d — размерность пространства.

Для моделирования процесса рекомбинации, описанного системой уравнений Смолуховского (1), в [12] представлен алгоритм Монте-Карло. На основе этого метода в данной работе разработана клеточно-автоматная модель рекомбинации электронов и дырок.

2. Клеточно-автоматная модель рекомбинации электронов и дырок в полупроводнике

Клеточный автомат (КА) — это дискретная динамическая система, состоящая из множества клеток, плотно заполняющих d -мерное пространство X [13–15]. Каждая клетка характеризуется парой значений (a, \mathbf{x}) , где a — это состояние клетки, \mathbf{x} — координата клетки в пространстве X . Состояние клетки $a \in A$, где A — алфавит состояний, который определяется возможными состояниями моделируемой системы. Состояние клетки изменяется в соответствии с правилами переходов Θ в зависимости от состояний самой клетки и взаимодействующих с ней клеток, которые выбираются с помощью шаблона моделирования T .

Шаблон моделирования может быть фиксированным и определяться как множество соседних клеток, расположенных вокруг центральной клетки, либо это может быть множество случайных клеток, выбираемых в соответствии с заданным распределением из множества всех клеток пространства. Применение правил переходов $\Theta(\mathbf{x})$ ко всем клеткам $\mathbf{x} \in X$ называется итерацией. Правила переходов могут применяться к клеткам $\mathbf{x} \in X$ в различном порядке. Этот порядок называется режимом функционирования (μ) КА. Для моделирования стохастических физико-химических процессов используется асинхронный режим работы КА ($\mu = \alpha$), при котором правила переходов применяются к случайно выбранным клеткам клеточного массива, сразу же изменяя их состояния, таким образом, новые состояния клеток вычисляются от состояний, полученных на предыдущих и на текущей итерациях.

На основании определения клеточного автомата [14] КА-модель рекомбинации электронов и дырок в полупроводнике задается в следующем виде: $\aleph = \langle A, X^d, \Theta, \alpha \rangle$. В соответствии с математической моделью рекомбинации (1) и алгоритмом Монте-Карло [12] алфавит состояний выбирается в виде $A = \{n, p, N_n, N_p, \emptyset\}$, где символ n обозначает электрон, p — дырку, N_n — рекомбинационный центр свободный для электрона, N_p — рекомбинационный центр свободный для дырки, \emptyset обозначает свободное место. Множество координат X^d в двумерном случае ($d = 2$) представимо в виде квадратной решетки $X^2 = \{\mathbf{x} = (i, j), i = 1 \dots Size_i, j = 1 \dots Size_j\}$, а в трехмерном ($d = 3$) — в виде куба $X^3 = \{\mathbf{x} = (i, j, k), i = 1 \dots Size_i, j = 1 \dots Size_j, k = 1 \dots Size_k\}$. На множестве координат вводятся периодические граничные условия.

Правила переходов задаются в следующем виде: $\Theta = R\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$, где символ R обозначает вероятностный выбор одного из правил θ_l , $l = 1, \dots, 5$. Правила переходов Θ моделируют процессы, происходящие в полупроводнике: θ_1 — радиационную рекомбинацию электрона и дырки, θ_2 — захват электрона в свободный рекомбинационный центр, θ_3 — захват дырки в рекомбинационный центр, содержащий электрон, θ_4 и θ_5 — диффузию электронов и дырок. Формально правила переходов записываются в виде (5):

$$\begin{aligned} \theta_1(\mathbf{x}) &: \{(n, \mathbf{x}), (p, \varphi_p(\mathbf{x}))\} \xrightarrow{p_1 \cdot \omega_1} \{(\emptyset, \mathbf{x}), (\emptyset, \varphi_p(\mathbf{x}))\}, \\ \theta_2(\mathbf{x}) &: \{(n, \mathbf{x}), (N_n, \varphi_{N_n}(\mathbf{x}))\} \xrightarrow{p_2 \cdot \omega_2} \{(\emptyset, \mathbf{x}), (N_p, \varphi_{N_n}(\mathbf{x}))\}, \\ \theta_3(\mathbf{x}) &: \{(p, \mathbf{x}), (N_p, \varphi_{N_p}(\mathbf{x}))\} \xrightarrow{p_3 \cdot \omega_3} \{(\emptyset, \mathbf{x}), (N_n, \varphi_{N_p}(\mathbf{x}))\}, \\ \theta_4(\mathbf{x}) &: \{(n, \mathbf{x}), (a, \psi(\mathbf{x}))\} \xrightarrow{p_4 \cdot \omega_4} \{(\emptyset, \mathbf{x}), (a', \psi(\mathbf{x}))\}, \\ \theta_5(\mathbf{x}) &: \{(p, \mathbf{x}), (b, \psi(\mathbf{x}))\} \xrightarrow{p_5 \cdot \omega_5} \{(\emptyset, \mathbf{x}), (b', \psi(\mathbf{x}))\}, \end{aligned} \quad (5)$$

$$a' = \begin{cases} n, & \text{если } a = \emptyset, \\ \emptyset, & \text{если } a = p, \\ N_p, & \text{если } a = N_n, \end{cases} \quad b' = \begin{cases} p, & \text{если } b = \emptyset, \\ \emptyset, & \text{если } b = n, \\ N_n, & \text{если } b = N_p, \end{cases}$$

где $\varphi_u(\mathbf{x})$, $u \in \{p, N_n, N_p\}$, и $\psi(\mathbf{x})$ — это координаты клеток, выбранных для взаимодействия с клеткой \mathbf{x} . Клетка $\varphi_u(\mathbf{x})$ выбирается случайным образом из множества всех клеток с состоянием u , присутствующих в клеточном массиве в данный момент времени t : $T_u(t) = \{(a, \mathbf{y}) : a = u, \mathbf{y} \in X\}$. Клетка $\psi(\mathbf{x})$ выбирается случайным образом из соседних клеток клетки \mathbf{x} по шаблону "крест", который в двумерном случае состоит из четырех клеток $T_4 = \{(i, j - 1), (i + 1, j), (i, j + 1), (i - 1, j)\}$, а в трехмерном случае — из шести клеток $T_6 = \{(i, j - 1, k), (i + 1, j, k), (i, j + 1, k), (i - 1, j, k), (i, j, k - 1), (i, j, k + 1)\}$.

При применении Θ выбирается одно из правил θ_l , $l = 1, \dots, 5$, с вероятностью p_l . Вероятности применения правил переходов вычисляются по формуле (6):

$$p_l = \lambda_l / \lambda, \quad l = 1, \dots, 5, \quad \lambda = \sum_{l=1}^5 \lambda_l, \quad (6)$$

$$\begin{aligned} \text{где } \lambda_1 &= C_n \cdot C_p \cdot B_0 \cdot \exp(-r_{np}^{min}/a_{np}), \quad \lambda_2 = C_n \cdot C_{N_n} \cdot b_{n0} \cdot \exp(-r_{nN_n}^{min}/a_{nN_n}), \\ \lambda_3 &= C_p \cdot C_{N_p} \cdot b_{p0} \cdot \exp(-r_{pN_p}^{min}/a_{pN_p}), \quad \lambda_4 = C_n \cdot \tilde{D}_n^d, \quad \text{где } \tilde{D}_n^2 = D_n, \quad \tilde{D}_n^3 = D_n \cdot r_0, \\ \lambda_5 &= C_p \cdot \tilde{D}_p^d, \quad \text{где } \tilde{D}_p^2 = D_p, \quad \tilde{D}_p^3 = D_p \cdot r_0, \end{aligned}$$

здесь $C_n, C_p, C_{N_n}, C_{N_p}$ — это количество электронов, дырок, рекомбинационных центров для электронов и рекомбинационных центров для дырок в клеточном массиве. B_0 — это коэффициент радиационной рекомбинации электронов и дырок, b_{n0} — коэффициент захвата электронов в свободные рекомбинационные центры, b_{p0} — коэффициент захвата дырок в рекомбинационные центры, содержащие электроны. D_n и D_p — это коэффициенты диффузии электронов и дырок, r_0 — радиус частицы. r_{uv}^{min} — это минимальное расстояние между всеми частицами типа u и v , где $u \in \{n, p\}$, $v \in \{p, N_n, N_p\}$. Аналогично, a_{uv} — это характерное расстояние взаимодействия частиц типа u с частицами типа v , где $u \in \{n, p\}$, $v \in \{p, N_n, N_p\}$.

После выбора одного из правил θ_l с соответствующей вероятностью p_l случайным образом выбирается клетка $\mathbf{x} \in X^d$ и взаимодействующая с ней клетка $\varphi_u(\mathbf{x})$ либо $\psi(\mathbf{x})$. В выбранных клетках правила переходов θ_l , $l = 1, 2, 3$, реализуются с вероятностью $\omega_l = \exp((r_{uv}^{min} - r_{uv})/a_{uv})$, где u, v соответствуют типам частиц в выбранном правиле θ_l . Вероятности реализации правил θ_4 и θ_5 : $\omega_4 = \omega_5 = 1$.

Каждой попытке применения одного из правил переходов θ_l , $l = 1, \dots, 3$, соответствует локальный временной шаг $\Delta\tau_{tun} = -\ln(rand_1)/\lambda$, где $rand_1 \in (0, 1)$ — случайное число. Попытке применения правил θ_4, θ_5 , моделирующих диффузию частиц, соответствует локальный временной шаг $\Delta\tau_{diff} = 1/(2D_u \cdot C_u)$, где $u \in \{n, p\}$, D_u — это коэффициент диффузии, соответствующий выбранному правилу, C_u — количество частиц, для которых моделируется диффузия. Кроме локального временного шага, в модели используется глобальный временной шаг $\Delta t = t_0 \cdot q^k$, где t_0 — это начальное время в секундах, q — коэффициент, отвечающий за длину глобального шага t_k , k — номер глобального временного шага. В данной модели глобальный временной шаг соответствует одной итерации КА.

Основными характеристиками, измеряемыми экспериментально, являются плотности частиц и интенсивность фотолюминесценции. В КА-модели рекомбинации \aleph плотность частиц типа u вычисляется как отношение количества клеток C_u с состоянием u в клеточном массиве в момент времени t_k к размеру клеточного массива $|X^d|$:

$$\rho_u(t_k) = C_u(t_k)/|X^d|, \quad u \in \{n, p, N_n, N_p\}. \quad (7)$$

Интенсивность фотолюминесценции вычисляется по формуле (8):

$$I(t_k) = \sigma/(t_k - t_{k-1}), \quad (8)$$

где σ — число фотонов, то есть число взаимодействий электронов и дырок, произошедшее за время $(t_k - t_{k-1})$.

В начальный момент времени электроны, дырки и рекомбинационные центры случайно и равномерно распределены в клеточном массиве. В связи с тем, что правила переходов и начальные данные КА-модели \aleph являются вероятностными, значения характеристик $\rho_u(t_k)$, $u \in \{n, p, N_n, N_p\}$, и $I(t_k)$ являются случайными величинами, поэтому на основе закона больших чисел вычисляются оценки этих характеристик как математические ожидания значений характеристик, полученных в результате КА-моделирования рекомбинации для различных начальных распределений частиц. Для вычисления оценок характеристик с высокой точностью необходимо, чтобы объем выборки был достаточно большим ($10^5 - 10^6$), для этого требуется провести большое количество численных экспериментов. В зависимости от размеров задачи время вычислений составляет от нескольких часов до нескольких суток. Например, КА-моделирование рекомбинации для двумерной области размером 1000×1000 клеток занимает 27,7 часов. Параллельная реализация задачи дает возможность существенно сократить время вычислений.

3. Параллельная реализация клеточно-автоматной модели рекомбинации электронов и дырок в полупроводнике

Традиционный метод распараллеливания клеточных автоматов — разбиение области на подобласти и распределение этих подобластей между процессами, для КА-модели рекомбинации оказывается малоэффективным в связи с тем, что в данной модели парами для взаимодействия являются не соседние клетки, а любые две клетки массива. Следовательно, при каждом вычислении новых состояний клеток \mathbf{x} и $\varphi_u(\mathbf{x})$ необходимо убедиться, что их состояния не были изменены другим процессом и обеспечить передачу новых состояний клеток при использовании их другими процессами. Это требует значительных накладных расходов на обмен данными.

Более эффективным для КА-модели рекомбинации является метод распределения вычислительных экспериментов с различными начальными данными между процессами. В работе для параллельной реализации КА-модели \aleph используется технология OpenMP. Множество вычислительных экспериментов распределяется между доступными потоками. Каждый поток генерирует свою последовательность случайных чисел, которая используется для начального распределения частиц в клеточном массиве, а также для выбора и применения правил переходов Θ . Все потоки независимо вычисляют значения характеристик на временных шагах $[t_0; t_{fin}]$, затем полученные значения суммируются главным потоком.

Эффективность параллельной реализации КА-модели \aleph проанализирована на примере задачи рекомбинации электронов и дырок при наличии рекомбинационных центров и диффузии электронов: $B_0 = 0,04 \text{ нс}^{-1}$, $b_{n0} = b_{p0} = 0,02 \text{ нс}^{-1}$, $a_{np} = 4 \text{ нм}$, $a_{nN_n} = a_{pN_p} = 2 \text{ нм}$, $t_0 = 0,5 \text{ нс}$, $t_{fin} = 10^8 \text{ нс}$, $C_n(0) = C_p(0) = 10000$, $C_{N_n}(0) = 5000$, $C_{N_p}(0) = 0$, $D_n = 1 \text{ нм}^2 \cdot \text{нс}^{-1}$, $D_p = 0 \text{ нм}^2 \cdot \text{нс}^{-1}$, $|X^d| = 10^6$ клеток. Число численных экспериментов с различными начальными данными, по которым оценивались значения характеристик, выбрано равным 1024. Вычисления проводились на кластере НКС—30Т Сибирского Суперкомпьютерного Центра СО РАН.

Для оценки качества параллельной реализации КА-модели рекомбинации вычисляются ускорение $S(th) = T(1)/T(th)$ и эффективность распараллеливания $Q(th) = T(1)/(T(th) \cdot th)$, где $T(th)$ — время вычислений при использовании th потоков. Значения $S(th)$ и $Q(th)$, полученные в результате параллельной реализации КА-модели \aleph в двумерном и трехмерном случаях представлены на рис. 1.

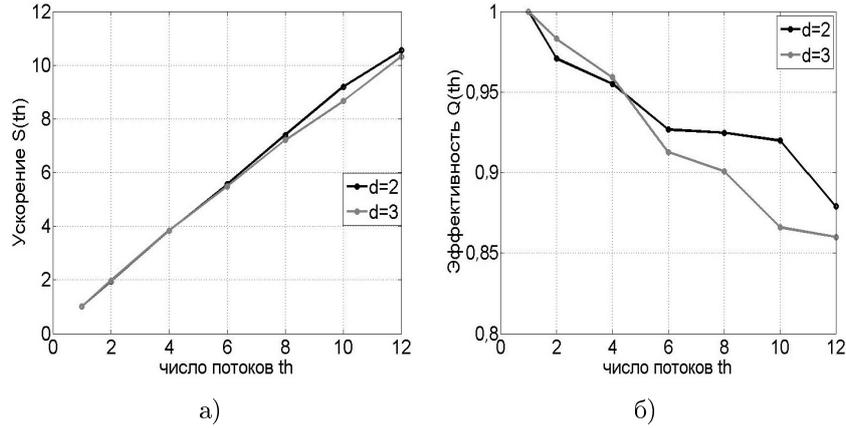


Рис. 1. Оценки качества параллельной реализации КА-модели рекомбинации: а) ускорение и б) эффективность распараллеливания

Ускорение и эффективность распараллеливания программы, реализующей двумерную КА-модель рекомбинации, выше, чем программы, реализующей трехмерную КА-модель \aleph . На рис. 1 видно, что ускорение $S(th)$ в обоих случаях ($d = 2$ и $d = 3$) близко к линейному, эффективность распараллеливания $Q(th)$ при использовании 12 потоков выше 0,85.

4. Результаты клеточно-автоматного моделирования рекомбинации электронов и дырок в полупроводнике

Основными параметрами КА-модели \aleph , влияющими на динамику процесса рекомбинации электронов и дырок, являются: константы скорости взаимодействия частиц (B_0, b_n, b_p), характерные расстояния взаимодействия ($a_{np}, a_{nN_n}, a_{pN_p}$), коэффициенты диффузии (D_n, D_p), начальное количество частиц ($C_n(0), C_p(0), C_{N_n}(0), C_{N_p}(0)$), и размер области ($Size_i, Size_j, Size_k$).

В ходе КА-моделирования на каждом глобальном шаге t_k вычисляются плотности частиц $\rho_u(t_k), u \in \{n, p, N_n, N_p\}$, (7) и интенсивность фотолюминесценции $I(t_k)$ (8).

С помощью КА-моделирования исследовалась динамика рекомбинации электронов и дырок в двумерной и трехмерной области для следующих режимов протекания процесса:

1. Радиационная рекомбинация электронов и дырок вследствие чистого туннелирования.
2. Радиационная рекомбинация электронов и дырок вследствие чистой диффузии частиц.
3. Радиационная и нерадиационная рекомбинация электронов и дырок.
4. Радиационная и нерадиационная рекомбинация частиц при наличии диффузии.

Рассмотрим динамику процесса рекомбинации электронов и дырок на примере КА-модели \aleph со следующими значениями параметров: $B_0 = 0,04 \text{ нс}^{-1}$, $b_{n0} = b_{p0} = 0,02 \text{ нс}^{-1}$, $a_{np} = 4 \text{ нм}$, $a_{nN_n} = a_{pN_p} = 2 \text{ нм}$, $D_n = D_p = 0,5 \text{ нм}^2 \cdot \text{нс}^{-1}$, $t_0 = 0,5 \text{ нс}$, $\rho_n(0) = \rho_p(0) = 0,01$, $\rho_{N_n}(0) = 0,00025$, $\rho_{N_p}(0) = 0$, $Size_i = Size_j = Size_k = 200$ клеток. В двумерном случае начальное количество частиц: $C_n(0) = C_p(0) = 400$, $C_{N_n}(0) = 10$, $C_{N_p}(0) = 0$, в трехмерном случае $C_n(0) = C_p(0) = 80000$, $C_{N_n}(0) = 2000$, $C_{N_p}(0) = 0$.

В режиме чистого туннелирования (режим 1) отсутствуют рекомбинационные центры и диффузия частиц: $C_{N_n}(0) = C_{N_p}(0) = 0$, $b_{n0} = b_{p0} = 0 \text{ нс}^{-1}$ и $D_n = D_p = 0 \text{ нм}^2 \cdot \text{нс}^{-1}$. Эволюция \aleph , имитирующая динамику процесса рекомбинации, для приведенных выше значений параметров, представлена на рис. 2. В начальный момент времени электроны и дырки равномерно распределены в объеме с плотностью $\rho_n(0) = \rho_p(0) = 0,01$ (рис. 2 а, г). В связи с тем, что в трехмерной области в начале моделирования число частиц очень ве-

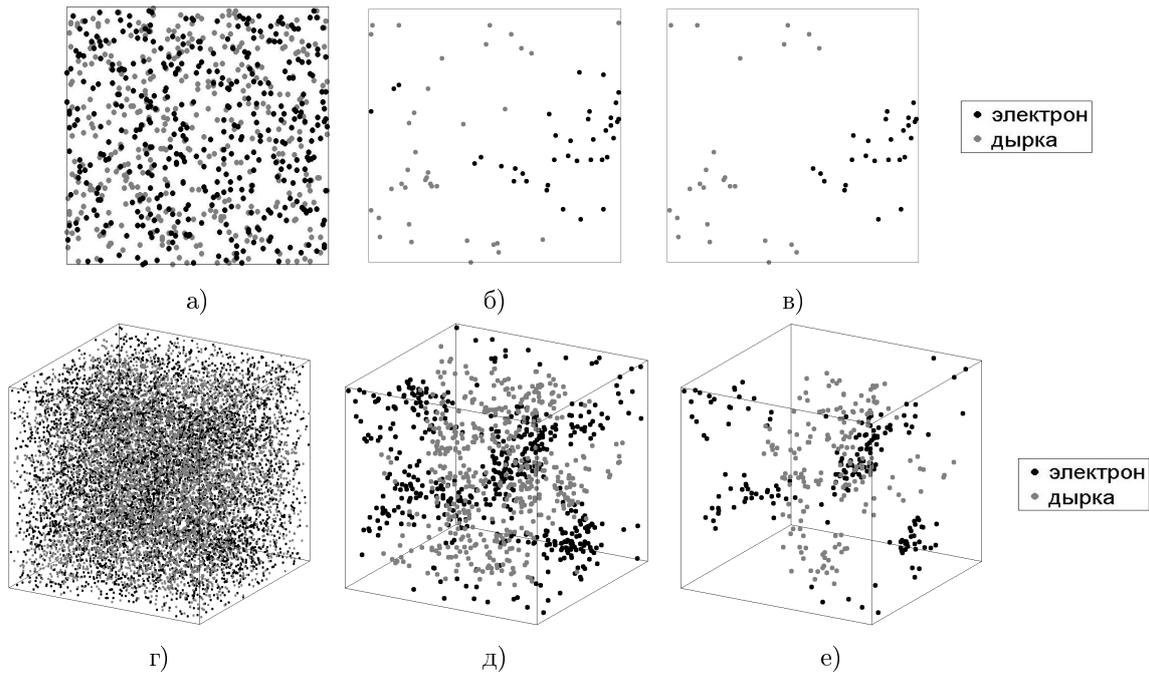


Рис. 2. Пространственное распределение частиц в режиме чистого туннелирования для $d = 2$: а) равномерное распределение при $t_0 = 0$ нс, б) формирование кластеров электронов и дырок при $t = 10^4$ нс, в) формирование макрокластеров при $t = 1,2 \cdot 10^7$ нс; и для $d = 3$: г) равномерное распределение при $t_0 = 21,5$ нс, д) формирование кластеров электронов и дырок при $t = 10^4$ нс, е) формирование макрокластеров при $t = 1,2 \cdot 10^7$ нс

лико ($C_n(0) = C_p(0) = 80000$), для наглядности на рисунке представлено распределение частиц в момент времени $t_0 = 21,5$ нс. Чем ближе расположены электрон и дырка, тем с большей вероятностью они взаимодействуют с друг другом. Следовательно, электроны и дырки, расположенные на близком расстоянии, достаточно быстро аннигилируют, в результате происходит пространственное разделение электронов и дырок (рис. 2 б, д). Далее происходит взаимодействие электронов и дырок, расположенных на границах кластеров, в результате мелкие кластеры исчезают и формируются макрокластеры (рис. 2 в, е).

Значения плотности электронов и интенсивности фотолюминесценции, полученные при КА-моделировании чистого туннелирования в двумерном и трехмерном случаях, представлены на рис. 3 в логарифмической шкале по обеим осям. Параллельность графиков асимптотического поведения характеристик $\rho'_{2D}(t)$, $\rho'_{3D}(t)$, $I'_{2D}(t)$, $I'_{3D}(t)$, вычисленных по формуле (3), и графиков характеристик $\rho_{2D}(t)$, $\rho_{3D}(t)$, $I_{2D}(t)$, $I_{3D}(t)$, полученных с помощью компьютерного моделирования, свидетельствует о соответствии теоретических и модельных результатов. Отклонение графиков плотности электронов $\rho_{2D}(t)$ и $\rho_{3D}(t)$ от асимптотических значений при больших временах $t > 10^{10}$ нс связано с малым количеством частиц, оставшихся в области. В начале моделирования, когда плотность частиц высокая, скорость аннигиляции электронов и дырок выше в трехмерной области, так как при одинаковой концентрации частиц направлений для поиска электронно-дырочных пар больше в трехмерной области, чем в двумерной. Следовательно, при применении правила перехода θ_1 , моделирующего радиационную рекомбинацию электрона и дырки вследствие туннелирования, при $d = 3$ пара для взаимодействия будет найдена с большей вероятностью, чем при $d = 2$. Далее, при $t > 10^3$ нс, графики характеристик, полученные с помощью КА-моделирования рекомбинации в двумерной и трехмерной области, становятся параллельны друг другу, что свидетельствует об одинаковой скорости аннигиляции частиц.

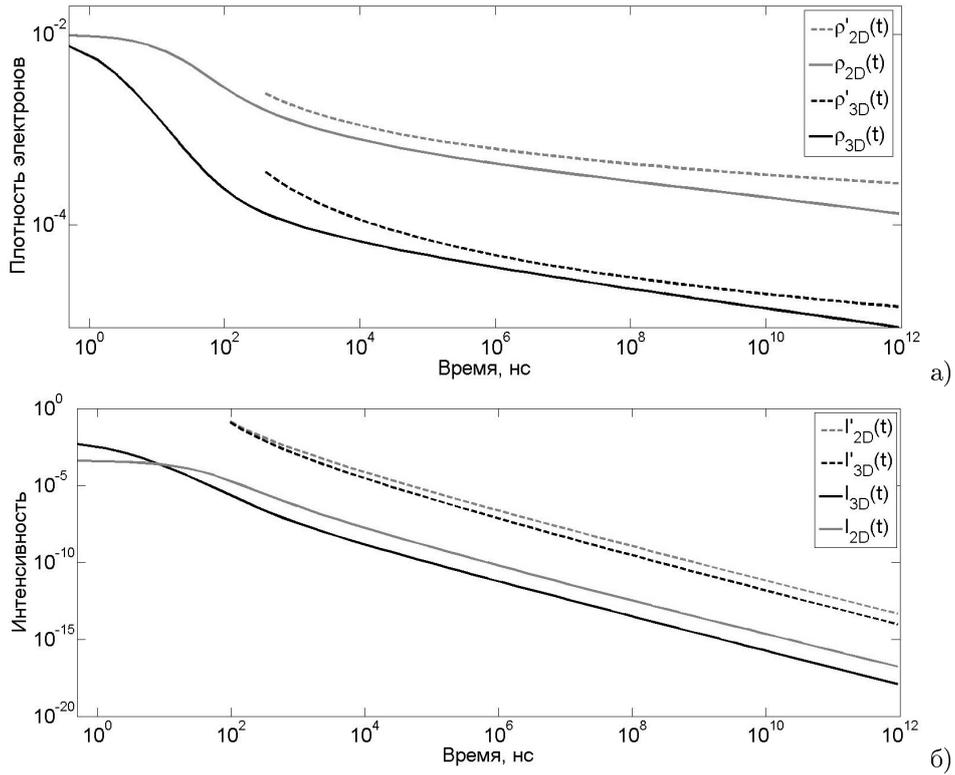


Рис. 3. Значения характеристик, полученные при КА-моделировании чистого туннелирования: а) плотность электронов, б) интенсивность фотолуминесценции

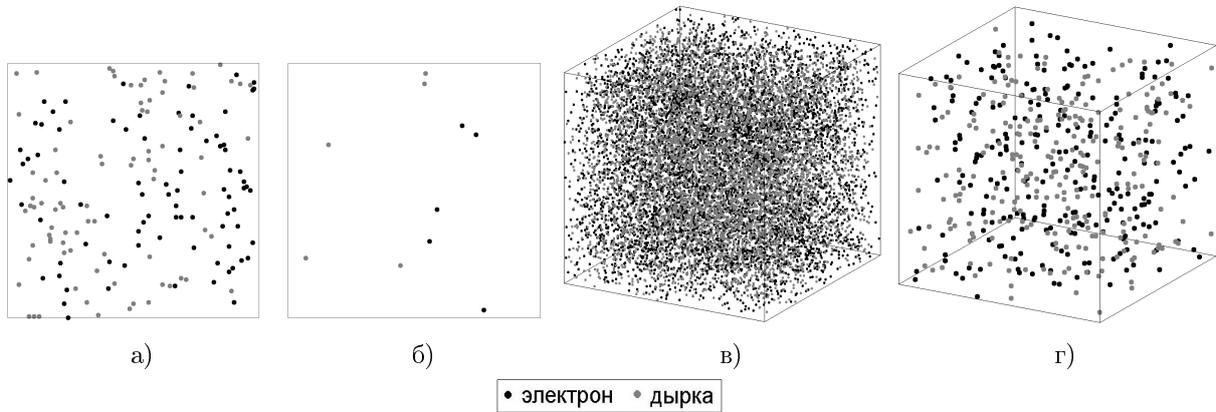


Рис. 4. Пространственное распределение электронов и дырок в режиме чистой диффузии: а) $d = 2, t = 600$ нс, б) $d = 2, t = 1,6 \cdot 10^4$ нс; в) $d = 3, t = 600$ нс, г) $d = 3, t = 1,6 \cdot 10^4$ нс

В режиме чистой диффузии (режим 2) туннелирование и рекомбинационные центры отсутствуют: $B_0 = 0 \text{ нс}^{-1}$, $C_{N_n}(0) = C_{N_p}(0) = 0$, $b_{n0} = b_{p0} = 0 \text{ нс}^{-1}$, коэффициенты диффузии $D_n = D_p = 0,5 \text{ нм}^2 \cdot \text{нс}^{-1}$. На рис. 4 показан характер эволюции \mathcal{N} , моделирующей динамику процесса рекомбинации вследствие диффузии. В этом случае рекомбинация электронов и дырок происходит при непосредственном столкновении частиц в одной клетке. В отличие от режима чистого туннелирования, из-за постоянного перемешивания частиц вследствие диффузии явного пространственного разделения электронов и дырок не происходит.

Значения характеристик, полученные в результате КА-моделирования чистой диффузии в двумерном и трехмерном случаях, представлены на рис. 5. В двумерном случае графики $\rho_{2D}(t)$ и $I_{2D}(t)$, вычисленные с помощью компьютерного моделирования, достаточно быстро отклоняются от графиков асимптотического поведения $\rho'_{2D}(t)$, $I'_{2D}(t)$, посчитанных

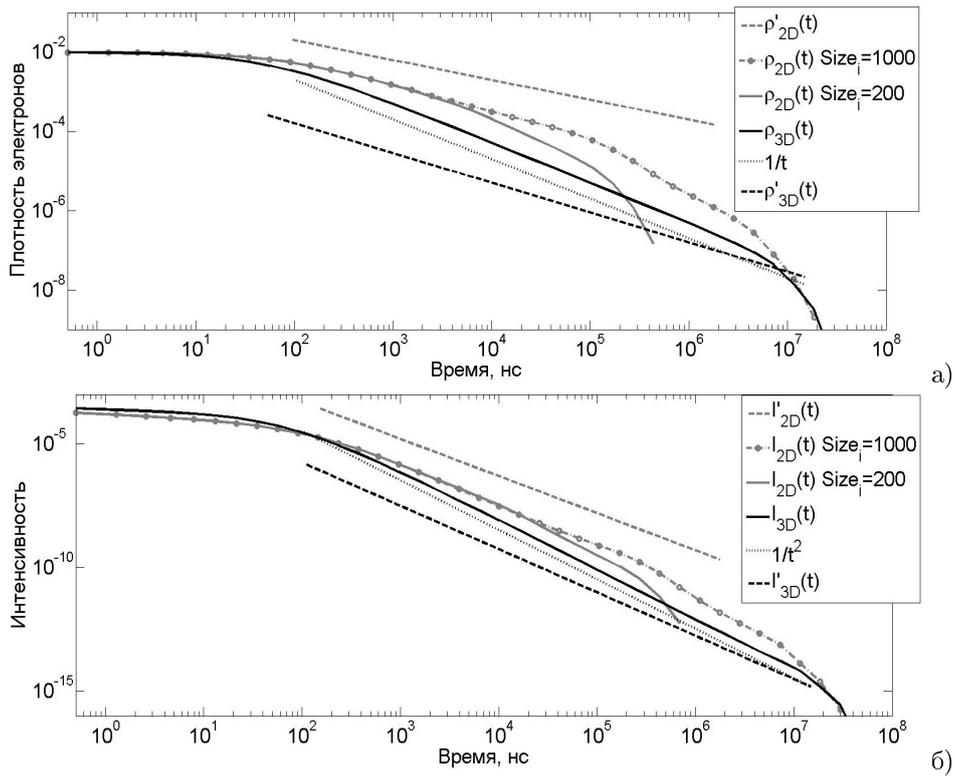


Рис. 5. Значения характеристик, полученные при КА-моделировании радиационной рекомбинации вследствие чистой диффузии частиц в двумерном и трехмерном случаях: а) плотность электронов, б) интенсивность фотолуминесценции

по формуле (4) с $d = 2$. Однако при увеличении размера области до 1000×1000 клеток и той же начальной плотности частиц при соответствующем увеличении количества электронов и дырок $C_n(0) = C_p(0) = 10000$ графики характеристик $\rho_{2D}(t)$ и $I_{2D}(t)$ параллельны графикам асимптотических значений $\rho'_{2D}(t)$, $I'_{2D}(t)$ до $t = 10^5$ нс, что подтверждает соответствие поведения КА-модели рекомбинации теоретическим данным. Тогда как в трехмерном случае при любом размере области графики $\rho_{3D}(t)$ и $I_{3D}(t)$ отличаются от асимптотического поведения $\rho'_{3D}(t)$, $I'_{3D}(t)$ (4), и хорошо согласуются с кривыми $\tilde{\rho}_{3D}(t) = 1/t$, $\tilde{I}_{3D}(t) = 1/t^2$, которые являются асимптотическими значениями характеристик для случая равномерного перемешивания частиц в объеме. Полученный результат показывает, что диффузия в трехмерном случае приводит к равномерному однородному распределению частиц во всем объеме. Кроме того, в отличие от режима чистого туннелирования, в трехмерном пространстве скорость аннигиляции частиц ниже, чем в двумерном.

Характер эволюции \aleph для режима радиационной и нерадиационной рекомбинации (режим 3) в присутствии рекомбинационных центров (для $d = 2$: $C_{N_n}(0) = 10$, для $d = 3$: $C_{N_n}(0) = 2000$) и без диффузии частиц показан на рис. 6. В присутствии рекомбинационных центров так же, как и в режиме чистого туннелирования, в результате аннигиляции электронов и дырок, расположенных на близком расстоянии, происходит формирование кластеров частиц. Наличие рекомбинационных центров приводит к увеличению скорости аннигиляции электронов и дырок, причем в трехмерном случае введение центров рекомбинации более существенно влияет на процесс, чем в двумерном случае.

Графики значений плотности электронов и интенсивности фотолуминесценции, полученные в результате КА-моделирования рекомбинации при наличии рекомбинационных центров для двумерного и трехмерного случаев, представлены на рис. 7. При сравнении рис.

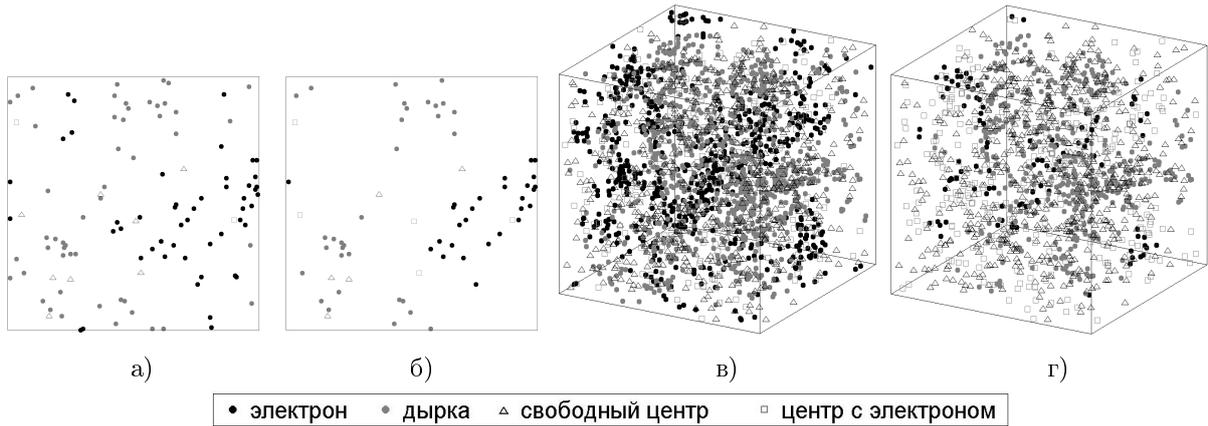


Рис. 6. Пространственное распределение электронов и дырок в режиме радиационной и нерадиационной рекомбинации: а) $d = 2$, $t = 960$ нс, б) $d = 2$, $t = 1,06 \cdot 10^5$ нс; в) $d = 3$, $t = 960$ нс, г) $d = 3$, $t = 1,06 \cdot 10^5$ нс

7 и рис. 3 видно, что при добавлении рекомбинационных центров значения характеристик $\rho_{2D}(t)$ и $I_{2D}(t)$ отклонились от значений, вычисленных для режима чистого туннелирования, меньше, чем значения $\rho_{3D}(t)$ и $I_{3D}(t)$. Следовательно, наличие рекомбинационных центров оказывает большее влияние на скорость рекомбинации в трехмерном пространстве.

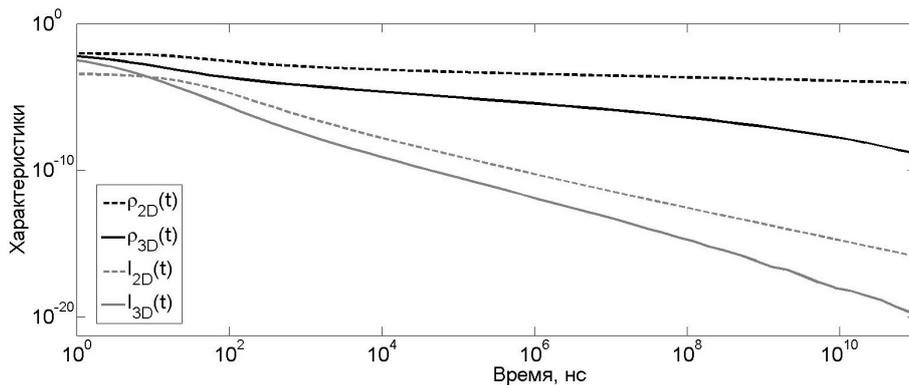


Рис. 7. Значения характеристик, полученные при КА-моделировании рекомбинации при наличии рекомбинационных центров в двумерном и трехмерном случаях

Характер эволюции \aleph в случае радиационной и нерадиационной рекомбинации (режим 4) при наличии рекомбинационных центров, распределенных в области с плотностью $\rho_{N_n}(0) = 0,00025$, $\rho_{N_p}(0) = 0$ при $b_{n0} = b_{p0} = 0,02$ нс⁻¹, и диффузии частиц $D_n = D_p = 0,5$ нм² · нс⁻¹ представлен на рис. 8. В результате аннигиляции под воздействием трех процессов: туннелирования, рекомбинации в центрах и диффузии, плотность частиц быстро уменьшается. Скорость туннелирования и рекомбинации в ловушках выше скорости диффузии, поэтому на каждом временном шаге можно выделить кластеры электронов и дырок, но вследствие перемешивания частиц эти кластеры быстро распадаются.

На рис. 9 представлены графики характеристик, полученные в результате КА-моделирования радиационной и нерадиационной рекомбинации при наличии рекомбинационных центров и диффузии частиц (режим 4). Аналогично режимам 1, 3, графики плотности электронов и интенсивности фотolumинесценции, вычисленные для трехмерной области, убывают быстрее, чем графики $\rho_{2D}(t)$ и $I_{2D}(t)$, полученные для двумерной области.

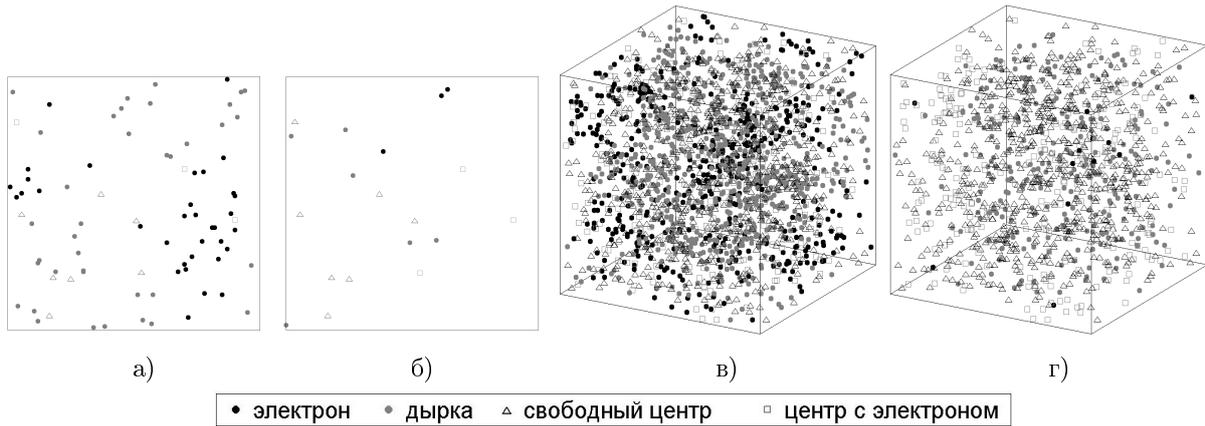


Рис. 8. Пространственное распределение электронов и дырок в режиме радиационной и нерадиационной рекомбинации при наличии диффузии: а) $d = 2$, $t = 960$ нс, б) $d = 2$, $t = 10^4$ нс; в) $d = 3$, $t = 960$ нс, г) $d = 3$, $t = 10^4$ нс

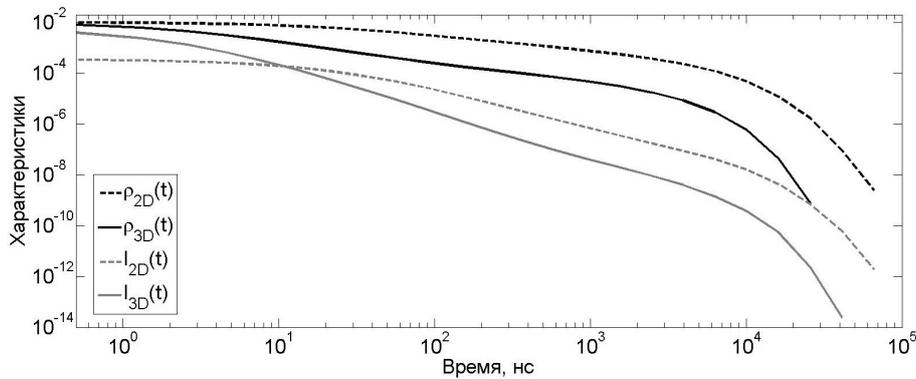


Рис. 9. Значения характеристик, полученные при КА-моделировании рекомбинации при наличии рекомбинационных центров и диффузии частиц для $d = 2$ и $d = 3$

Заклучение

На основе системы уравнений Смолуховского (1) и алгоритма Монте-Карло [9, 12] разработана КА-модель рекомбинации электронов и дырок в двумерном и трехмерном полупроводниках. Для моделирования областей больших размеров и вычисления интегральных характеристик процесса рекомбинации (плотности частиц и интенсивности фотолюминесценции) разработаны параллельные программы, реализующие КА-модель рекомбинации в двумерном и трехмерном случаях. Параллельные программы реализованы с помощью технологии OpenMP. Эффективность распараллеливания на кластере НКС–30Т Сибирского Суперкомпьютерного Центра СО РАН выше 85% при использовании 12 потоков.

С помощью разработанной параллельной программы изучена кинетика процесса рекомбинации для четырех режимов: 1) радиационной рекомбинации электронов и дырок вследствие чистого туннелирования, 2) радиационной рекомбинации электронов и дырок вследствие чистой диффузии, 3) радиационной и нерадиационной рекомбинации при наличии центров рекомбинации, 4) радиационной и нерадиационной рекомбинации при наличии центров рекомбинации и диффузии частиц. В результате исследования эволюции КА-модели обнаружено формирование кластеров электронов и дырок при малых значениях коэффициентов диффузии частиц. Выявлено, что в режимах, в которых присутствует туннелирование частиц, (режимы 1, 3, 4) скорость аннигиляции электронов и дырок в трехмерном случае выше, чем в двумерном. Кроме того, наличие центров рекомбинации

оказывает большее влияние на процесс рекомбинации в трехмерном пространстве. Однако в режиме чистой диффузии в двумерной области аннигиляция электронов и дырок происходит быстрее, чем в трехмерной, что свидетельствует о более существенном влиянии диффузии частиц на скорость рекомбинации в двумерном случае.

Работа выполнена при финансовой поддержке Российского научного фонда, грант № 14-11-00083.

Литература

1. Туркин А. Нитрид галлия как один из перспективных материалов в современной оптоэлектронике // Компоненты и технологии. 2011. No 5. С. 6–10.
2. Грузинцев А.Н., Редькин А.Н., Barthou С. Излучательная рекомбинация нанокристаллов GaN при большой мощности оптического возбуждения // Физика и техника полупроводников. 2005. No 10. С. 1200–1203.
3. Gorgis A., Flissikowski T., Brandt O., Cheze C., Geelhaar L., Riechert H., and Grahn H.T. Time-resolved photoluminescence spectroscopy of individual GaN nanowires // Physical review B. 2012. Vol. 86. 041302(R). DOI: 10.1103/physrevb.86.041302.
4. Hong G., Guoqiang R., Taofei Z., Feifei T., Yu X., Yumin Z., Mingyue W., Zhiqiang Z., Demin C., Jianfeng W., Ke X. Study of optical properties of bulk GaN crystals grown by HVPE // Journal of Alloys and Compounds. 2016. Vol. 674. P. 218–222.
5. Bulashevich K.A., Mymrin V.F., Karpov S.Y., Zhmakin I.A., Zhmakin A.I. Simulation of visible and ultra-violet group-III nitride light emitting diodes // Journal of Computational Physics. 2006, Vol. 213, Issue 1. P. 214–238. DOI: 10.1016/j.jcp.2005.08.011.
6. Massimo V.F., Steven E.L. Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects // Physical Review B. 1988. Vol. 38, No 14. P. 9721–9745. DOI: 10.1103/physrevb.38.9721.
7. Kotomin E., Kuzovkov V. Modern Aspects of Diffusion-Controlled Reactions. Volume 34: Cooperative Phenomena in Bimolecular Processes (Comprehensive Chemical Kinetics) // Elsevier Science, 1996. 611 p. DOI: 10.1016/s0069-8040(96)x8001-4.
8. Gillespie D.T. A diffusional bimolecular propensity function // Journal of Chemical Physics. 2009. Vol. 131, Issue 16. P. 164109-1–164109-13. DOI: 10.1063/1.3253798.
9. Sabelfeld K.K., Brandt O., Kaganer V.M. Stochastic model for the fluctuation-limited reaction-diffusion kinetics in inhomogeneous media based on the nonlinear Smoluchowski equations // J. Math. Chem. 2015. Vol. 53, Issue 2. P. 651–669. DOI: 10.1007/s10910-014-0446-6.
10. Kolodko A.A. and Sabelfeld K.K. Stochastic Lagrangian model for spatially inhomogeneous Smoluchowski equation governing coagulating and diffusing particles // Monte Carlo Methods and Applications. 2001. Vol. 7, No. 3-4. P. 223–228. DOI: 10.1515/mcma.2001.7.3-4.223.
11. Kolodko A., Sabelfeld K. and Wagner W. A stochastic method for solving Smoluchowski's coagulation equation // Mathematics and Computers in Simulation. 1999. Vol. 49, No 1-2. P. 57–79. DOI: 10.1016/s0378-4754(99)00008-7.

12. Sabelfeld K.K., Levykin A.I., Kireeva A.E. // Stochastic simulation of fluctuation-induced reaction-diffusion kinetics governed by Smoluchowski equations, Monte Carlo Methods and Applications. 2015. Vol. 21, No 1. P. 33–48. DOI: 10.1515/mcma-2014-0012.
13. Toffoli T., Margolus N. Cellular Automata Machines: A New Environment for Modeling // USA: MIT Press, 1987. 259 p.
14. Бандман О.Л. Клеточно-автоматные модели пространственной динамики // Системная информатика. Методы и модели современного программирования. 2006. No 10. С. 59–113.
15. Bandman O.L. Mapping physical phenomena onto CA-models // AUTOMATA-2008. In: Adamatzky A., Alonso-Sanz R., Lawniczak A., Martinez G.J., Morita K., Worsch T. (eds.) Theory and Applications of Cellular Automata. Luniver Press, UK, 2008. P. 381–397.

Сабельфельд Карл Карлович, д.ф.-м.н., профессор, главный научный сотрудник, лаборатория стохастических задач, Институт вычислительной математики и математической геофизики СО РАН (Новосибирск, Российская Федерация)

Киреева Анастасия Евгеньевна, к.ф.-м.н., научный сотрудник, лаборатория стохастических задач, Институт вычислительной математики и математической геофизики СО РАН (Новосибирск, Российская Федерация)

DOI: 10.14529/cmse170106

PARALLEL IMPLEMENTATION OF STOCHASTIC CELLULAR AUTOMATA MODEL OF ELECTRON-HOLE RECOMBINATION IN 2D AND 3D HETEROGENEOUS SEMICONDUCTORS

© 2017 K.K. Sabelfeld, A.E. Kireeva

*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk,
Russian Federation (Academician Lavrentyev Avenue, Novosibirsk, 630090 Russia)*

E-mail: karl@osmf.sccc.ru, kireeva@ssd.sccc.ru

Received: 28.10.2016

Parallel programs implementing stochastic cellular automata (CA) model of electron-hole recombination in an inhomogeneous semiconductor for two- and three-dimensional cases are developed. The spatio-temporal distributions of particles are investigated by the CA simulation. Spatial separation of electrons and holes with clusters formation is found and analyzed. Parallel implementation of the CA model allows us to calculate integral characteristics of the recombination process (particle densities and radiative intensity) in acceptable time. Recombination kinetics in the vicinity of the recombination centers and diffusion in two- and three-dimensional space is investigated using the parallel program.

Keywords: electron-hole recombination, semiconductor, parallel implementation, stochastic cellular automaton, radiative intensity.

FOR CITATION

Sabelfeld K.K., Kireeva A.E. Parallel implementation of stochastic cellular automata model of electron-hole recombination in 2D and 3D heterogeneous semiconductors. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2017. vol. 6, no. 1. pp. 87–103. (in Russian) DOI: 10.14529/cmse170106.

References

1. Turkin A. Nitrid galliya kak odin iz perspektivnykh materialov v sovremennoy optoelektronike [Nitride Gallium as One of the Perspective Materials in Modern Optoelectronics]. *Komponenty i tekhnologii* [Components and Technologies], 2011. no. 5. pp. 6–10. (in Russian)
2. Gruzintsev A.N., Red'kin A.N., Barthou C. Izluchatel'naya rekombinatsiya nanokristallov GaN pri bol'shoy moshchnosti opticheskogo vzbuzhdeniya [Radiative recombination of GaN nanocrystals at high power optical excitation]. *Fizika i tekhnika poluprovodnikov* [Physics and Technique of Semiconductors], 2005. no. 10. pp. 1200–1203. (in Russian)
3. Gorgis A., Flissikowski T., Brandt O., Cheze C., Geelhaar L., Riechert H., and Grahn H.T. Time-resolved photoluminescence spectroscopy of individual GaN nanowires. *Physical review B*. 2012. Vol. 86. 041302(R). DOI: 10.1103/physrevb.86.041302.
4. Hong G., Guoqiang R., Taofei Z., Feifei T., Yu X., Yumin Z., Mingyue W., Zhiqiang Z., Demin C., Jianfeng W., Ke X. Study of optical properties of bulk GaN crystals grown by HVPE. *Journal of Alloys and Compounds*. 2016. Vol. 674. P. 218–222.
5. Bulashevich K.A., Mymrin V.F., Karpov S.Y., Zhmakin I.A., Zhmakin A.I. Simulation of visible and ultra-violet group-III nitride light emitting diodes. *Journal of Computational Physics*. 2006, Vol. 213, Issue 1. P. 214–238. DOI: 10.1016/j.jcp.2005.08.011
6. Massimo V.F., Steven E.L. Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects. *Physical Review B*. 1988. Vol. 38, No 14. P. 9721–9745. DOI: 10.1103/physrevb.38.9721.
7. Kotomin E., Kuzovkov V. Modern Aspects of Diffusion-Controlled Reactions. Volume 34: Cooperative Phenomena in Bimolecular Processes (Comprehensive Chemical Kinetics). Elsevier Science, 1996. 611 p. DOI: 10.1016/s0069-8040(96)x8001-4.
8. Gillespie D.T. A diffusional bimolecular propensity function. *Journal of Chemical Physics*. 2009. Vol. 131, Issue 16. P. 164109-1–164109-13. DOI: 10.1063/1.3253798.
9. Sabelfeld K.K., Brandt O., Kaganer V.M. Stochastic model for the fluctuation-limited reaction-diffusion kinetics in inhomogeneous media based on the nonlinear Smoluchowski equations. *J. Math. Chem.* 2015. vol. 53, Issue 2. pp. 651–669. DOI: 10.1007/s10910-014-0446-6.
10. Kolodko A.A. and Sabelfeld K.K. Stochastic Lagrangian model for spatially inhomogeneous Smoluchowski equation governing coagulating and diffusing particles. *Monte Carlo Methods and Applications*. 2001. vol. 7, no. 3-4. pp. 223–228. DOI: 10.1515/mcma.2001.7.3-4.223.
11. Kolodko A., Sabelfeld K. and Wagner W. A stochastic method for solving Smoluchowski's coagulation equation. *Mathematics and Computers in Simulation*. 1999. vol. 49, no 1-2. pp. 57–79. DOI: 10.1016/s0378-4754(99)00008-7
12. Sabelfeld K.K., Levykin A.I., Kireeva A.E Stochastic simulation of fluctuation-induced reaction-diffusion kinetics governed by Smoluchowski equations. *Monte Carlo Methods and Applications*. 2015. vol. 21, no 1. pp. 33–48. DOI: 10.1515/mcma-2014-0012.
13. Toffoli T., Margolus N. Cellular Automata Machines: A New Environment for Modeling. USA: MIT Press, 1987. 259 p.
14. Bandman O.L. Kletchno-avtomatnye modeli prostranstvennoy dinamiki [Cellular Automata Model of Spatial Dynamics]. *Sistemnaya informatika. Metody i modeli sovremennogo*

programmirovaniya [System Informatics. Methods and Models of Modern Programming], 2006. no 10. pp. 59–113. (in Russian)

15. Bandman O.L. Mapping physical phenomena onto CA-models. AUTOMATA-2008. In: Adamatzky A., Alonso-Sanz R., Lawniczak A., Martinez G.J., Morita K., Worsch T. (eds.) Theory and Applications of Cellular Automata. Luniver Press, UK, 2008. pp. 381–397.

СВЕДЕНИЯ ОБ ИЗДАНИИ

Научный журнал «Вестник ЮУрГУ. Серия «Вычислительная математика и информатика» основан в 2012 году.

Свидетельство о регистрации ПИ ФС77-57377 выдано 24 марта 2014 г. Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций.

Журнал включен в Реферативный журнал и Базы данных ВИНИТИ; индексируется в библиографической базе данных РИНЦ. Журнал размещен в открытом доступе на Всероссийском математическом портале MathNet. Сведения о журнале ежегодно публикуются в международной справочной системе по периодическим и продолжающимся изданиям «Ulrich's Periodicals Directory».

Решением Президиума Высшей аттестационной комиссии Министерства образования и науки Российской Федерации журнал включен в «Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук» по следующим отраслям и группам специальностей: 05.13.00 – информатика, вычислительная техника и управление; 01.01.00 – математика; 25.00.00 – науки о Земле (№421).

Подписной индекс научного журнала «Вестник ЮУрГУ», серия «Вычислительная математика и информатика»: 10244, каталог «Пресса России». Периодичность выхода — 4 выпуска в год (февраль, май, август и ноябрь).

ПРАВИЛА ДЛЯ АВТОРОВ

1. Правила подготовки рукописей и пример оформления статей можно загрузить с сайта серии <http://vestnikvmi.susu.ru>. **Статьи, оформленные без соблюдения правил, к рассмотрению не принимаются.**
2. Адрес редакции научного журнала «Вестник ЮУрГУ», серия «Вычислительная математика и информатика»:
Россия 454080, г. Челябинск, пр. им. В.И. Ленина, 76, ЮУрГУ, кафедра СП,
ответственному секретарю Цымблеру М.Л.
3. Адрес электронной почты редакции: vestnikvmi@susu.ru
4. **Плата с авторов за публикацию рукописей не взимается, и гонорары авторам не выплачиваются.**