

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

6(214)  
2014

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ  
Издательство "Новые технологии"

## СОДЕРЖАНИЕ

### ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

- Болотова Л. С., Карасев А. А., Старых В. А.** Формализация экспертных знаний для управления инцидентами информационных систем на основе онтологического подхода . . . . . 3
- Лебедев Г. Н., Семенов М. Е., Грачиков Д. В., Канищева О. И.** Гистерезисная модель синхронизации биологических нейронов . . . . . 11
- Сакулин С. А., Алфимцев А. Н.** Формализация экспертных знаний об удобстве веб-страниц на основе агрегирования пользовательских критериев . . . . . 16
- Кузьмин А. А., Адуенко А. А., Стрижов В. В.** Тематическая классификация тезисов крупной конференции с использованием экспертной модели . . . . . 22

### ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И СЕТИ

- Бобков С. Г., Аряшев С. И., Барских М. Е., Зубковский П. С., Ивасюк Е. В.** Высокпроизводительные расширения архитектуры универсальных микропроцессоров для ускорения инженерных расчетов . . . . . 27
- Шилов С. Н., Кургалин С. Д., Крыловецкий А. А.** Анализ и реализация механизма репликации ресурсных записей в DNS-кластере . . . . . 38

### МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

- Аристов А. О.** Особенности моделирования потоковых систем на основе квазиклассических сетей с использованием структурной методологии проектирования . . 44
- Димов Э. М., Маслов О. Н., Трошин Ю. В.** Снижение неопределенности выбора управленческих решений с помощью метода статистического имитационного моделирования . . . . . 51

### СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

- Гаврилов С. В., Рыжова Д. И., Стемпковский А. Л.** Проблема анализа пикового тока при проектировании сверхбольших интегральных схем на логическом уровне и современные методы ее решения . . . . . 58

### ГЕОИНФОРМАЦИОННЫЕ СИСТЕМЫ

- Цветков В. Я.** Ситуационное моделирование в геоинформатике . . . . . 64

### ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

- Юрин А. И., Красивская М. И., Дмитриев А. В., Злодеев Г. Ю.** Применение виртуальных лабораторных стендов в образовательном процессе . . . . . 70
- Приложение. Полунов Ю. Л., Шилов В. В.** БББ: Бэббидж. Биография. Библиография

**Главный редактор:**  
СТЕМПКОВСКИЙ А. Л.,  
акад. РАН, д. т. н., проф.

**Зам. главного редактора:**  
ИВАННИКОВ А. Д., д. т. н., проф.  
ФИЛИМОНОВ Н. Б., д. т. н., с.н.с.

**Редакционный совет:**  
БЫЧКОВ И. В., акад. РАН, д. т. н.  
ЖУРАВЛЕВ Ю. И.,  
акад. РАН, д. ф.-м. н., проф.  
КУЛЕШОВ А. П.,  
акад. РАН, д. т. н., проф.  
ПОПКОВ Ю. С.,  
чл.-корр. РАН, д. т. н., проф.  
РУСАКОВ С. Г.,  
чл.-корр. РАН, д. т. н., проф.  
СОЙФЕР В. А.,  
чл.-корр. РАН, д. т. н., проф.  
СОКОЛОВ И. А., акад.  
РАН, д. т. н., проф.  
СУЕТИН Н. В., д. ф.-м. н., проф.  
ЧАПЛЫГИН Ю. А.,  
чл.-корр. РАН, д. т. н., проф.  
ШАХНОВ В. А.,  
чл.-корр. РАН, д. т. н., проф.  
ШОКИН Ю. И.,  
акад. РАН, д. т. н., проф.  
ЮСУПОВ Р. М.,  
чл.-корр. РАН, д. т. н., проф.

**Редакционная коллегия:**  
АВДОШИН С. М., к. т. н., доц.  
АНТОНОВ Б. И.  
БАРСКИЙ А. Б., д. т. н., проф.  
ВАСЕНИН В. А., д. ф.-м. н., проф.  
ГАЛУШКИН А. И., д. т. н., проф.  
ДИМИТРИЕНКО Ю. И., д. ф.-м. н., проф.  
ДОМРАЧЕВ В. Г., д. т. н., проф.  
ЗАГИДУЛЛИН Р. Ш., к. т. н., доц.  
ЗАРУБИН В. С., д. т. н., проф.  
ИСАЕНКО Р. О., к. т. н., с.н.с.  
КАРПЕНКО А. П., д. ф.-м. н., проф.  
КОЛИН К. К., д. т. н., проф.  
КУЛАГИН В. П., д. т. н., проф.  
КУРЕЙЧИК В. М., д. т. н., проф.  
КУХАРЕНКО Б. Г., к. ф.-м. н., доц.  
ЛЬВОВИЧ Я. Е., д. т. н., проф.  
МИХАЙЛОВ Б. М., д. т. н., проф.  
НЕЧАЕВ В. В., к. т. н., проф.  
РЯБОВ Г. Г., чл.-корр. РАН, д. т. н., проф.  
СОКОЛОВ Б. В., д. т. н., проф.  
УСКОВ В. Л., к. т. н. (США)  
ФОМИЧЕВ В. А., д. т. н., проф.  
ЧЕРМОШЕНЦЕВ С. Ф., д. т. н., проф.  
ШИЛОВ В. В., к. т. н., доц.

**Редакция:**  
БЕЗМЕНОВА М. Ю.  
ГРИГОРИН-РЯБОВА Е. В.  
ЛЫСЕНКО А. В.  
ЧУГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://novtex.ru/IT>.  
Журнал включен в систему Российского индекса научного цитирования.  
Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

# INFORMATION TECHNOLOGIES

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

6(214)  
2014

THEORETICAL AND APPLIED SCIENTIFIC AND TECHNICAL JOURNAL

Published since November 1995

ISSN 1684-6400

## CONTENTS

### INTELLIGENT SYSTEMS AND TECHNOLOGIES

- Bolotova L. S., Karasev A. A., Starykh V. A.** Formalization of the Expert Knowledge for Incident's Management of Information System Based on the Ontological Approach . . . . . 3
- Lebedev G. N., Semenov M. E., Grachikov D. V., Kanishcheva O. I.** Hysteretic Synchronization Models of Biological Neurons . . . . . 11
- Sakulin S. A., Alfimtsev A. N.** Formalization of Expert Knowledge about the Usability of Web Pages Based on User Criteria Aggregation . . . . . 16
- Kuzmin A. A., Aduenko A. A., Strijov V. V.** Thematic Classification Using Expert Model for Major Conference Abstracts. . . . . 22

### COMPUTING SYSTEMS AND NETWORKS

- Bobkov S. G., Aryashev S. I., Barskyh M. E., Zubkovskiy P. S., Ivasyuk E. V.** High-Performance Extensions of Microprocessor Architecture for Speeding-Up of Scientific and Engineering Calculations . . . . . 27
- Shilov S. N., Kurgalin S. D., Krylovetsky A. A.** Analysis and Implementation of Resource Records Replication Mechanism within DNS Cluster . . . . . 38

### MODELING AND OPTIMIZATION

- Aristov A. O.** The Features of Quasi Cellular Nets in Simulations of Flow-Based Systems and Their Design with Structure Approach . . . . . 44
- Dimov A. M., Maslov O. N., Troshin Ju. V.** Reducing Uncertainty in a Choice of Management Decisions Using Statistical Simulation . . . . . 51

### CAD-SYSTEMS

- Gavrilov S. V., Ryzhova D. I., Stempkovsky A. L.** Problem of Peak Current Analysis at the Logic Level of VLSI Design and its Modern Solutions . . . . . 58

### GEOINFORMATION SYSTEMS

- Tsvetkov V. Ya.** Situational Modeling in Geoinformatics . . . . . 64

### INFORMATION TECHNOLOGIES IN EDUCATION

- Yurin A. I., Krasivskaya M. I., Dmitriev A. V., Zlodeev G. Yu.** Virtual Laboratory Stands Application in Education . . . . . 70

- APPLICATION. Polunov Yu. L., Shilov V. V.** BBB: Babbage. Biography. Bibliography

#### Editor-in-Chief:

Stempkovsky A. L., Member of RAS,  
Dr. Sci. (Tech.), Prof.

#### Deputy Editor-in-Chief:

Ivannikov A. D., Dr. Sci. (Tech.), Prof.  
Filimonov N. B., Dr. Sci. (Tech.), Prof.

#### Chairman:

Bychkov I. V., Member of RAS,  
Dr. Sci. (Tech.), Prof.  
Zhuravljov Yu. I., Member of RAS,  
Dr. Sci. (Phys.-Math.), Prof.  
Kuleshov A. P., Member of RAS,  
Dr. Sci. (Tech.), Prof.  
Popkov Yu. S., Corresp. Member of RAS,  
Dr. Sci. (Tech.), Prof.  
Rusakov S. G., Corresp. Member of RAS,  
Dr. Sci. (Tech.), Prof.  
Soifer V. A., Corresp. Member of RAS,  
Dr. Sci. (Tech.), Prof.  
Sokolov I. A., Member of RAS,  
Dr. Sci. (Phys.-Math.), Prof.  
Suetin N. V.,  
Dr. Sci. (Phys.-Math.), Prof.  
Chaplygin Yu. A., Corresp. Member of RAS,  
Dr. Sci. (Tech.), Prof.  
Shakhnov V. A., Corresp. Member of RAS,  
Dr. Sci. (Tech.), Prof.  
Shokin Yu. I., Member of RAS,  
Dr. Sci. (Tech.), Prof.  
Yusupov R. M., Corresp. Member of RAS,  
Dr. Sci. (Tech.), Prof.

#### Editorial Board Members:

Avdoshin S. M., Cand. Sci. (Tech.), Ass. Prof.  
Antonov B. I.  
Barsky A. B., Dr. Sci. (Tech.), Prof.  
Vasenin V. A., Dr. Sci. (Phys.-Math.), Prof.  
Galushkin A. I., Dr. Sci. (Tech.), Prof.  
Dimitrienko Yu. I., Dr. Sci. (Phys.-Math.), Prof.  
Domrachev V. G., Dr. Sci. (Tech.), Prof.  
Zagidullin R. Sh., Cand. Sci. (Tech.), Ass. Prof.  
Zarubin V. S., Dr. Sci. (Tech.), Prof.  
Isaenko R. O., Cand. Sci. (Tech.)  
Karpenko A. P., Dr. Sci. (Phys.-Math.), Prof.  
Kolin K. K., Dr. Sci. (Tech.)  
Kulagin V. P., Dr. Sci. (Tech.), Prof.  
Kureichik V. M., Dr. Sci. (Tech.), Prof.  
Kukharensko B. G., Cand. Sci. (Phys.-Math.)  
Ljvovich Ya. E., Dr. Sci. (Tech.), Prof.  
Mikhailov B. M., Dr. Sci. (Tech.), Prof.  
Nechaev V. V., Cand. Sci. (Tech.), Ass. Prof.  
Ryabov G. G., Corresp. Member of RAS,  
Dr. Sci. (Tech.), Prof.  
Sokolov B. V., Dr. Sci. (Tech.)  
Uskov V. L. (USA), Dr. Sci. (Tech.)  
Fomichev V. A., Dr. Sci. (Tech.), Prof.  
Chermoshentsev S. F., Dr. Sci. (Tech.), Prof.  
Shilov V. V., Cand. Sci. (Tech.), Ass. Prof.

#### Editors:

Bezmenova M. Yu.  
Grigorin-Ryabova E. V.  
Lysenko A. V.  
Chugunova A. V.

Complete Internet version of the journal at site: <http://novtex.ru/IT>.

According to the decision of the Higher Certifying Commission of the Ministry of Education of Russian Federation, the journal is inscribed in "The List of the Leading Scientific Journals and Editions wherein Main Scientific Results of Theses for Doctor's or Candidate's Degrees Should Be Published"

# ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ INTELLIGENT SYSTEMS AND TECHNOLOGIES

УДК 004.822+89

Л. С. Болотова<sup>1</sup>, д-р техн. наук, проф., А. А. Карасев, аспирант,  
В. А. Старых<sup>2</sup>, канд. техн. наук, проф., e-mail: starykh.vladimir@gmail.com  
<sup>1</sup> Национальный исследовательский университет "Высшая школа экономики"  
<sup>2</sup> ФГАУ ГНИИ ИТТ "Информика"

## Формализация экспертных знаний для управления инцидентами информационных систем на основе онтологического подхода

*Статья посвящена решению проблемы сбора и формализации экспертных знаний, представляющих инциденты, получаемых в процессе эксплуатации информационных систем и предназначенных для решения задачи повышения эффективности их управления. Рассматриваются существующие подходы и инструменты решения проблемы управления инцидентами, определяются их основные недостатки. В качестве перспективного подхода предлагается использование онтологии предметной области, обеспечивающей базис для формирования концептуальной модели предметной области управления информационными системами. Организация объектов предметной области осуществляется с использованием классификации инцидентов в информационных системах с использованием структурного подхода и классификатора, основанного на модели OSI. Изложены основные принципы разработки прототипа системы управления инцидентами, предназначенной для решения поставленных задач с использованием инструментальных средств Thinkmap SDK.*

**Ключевые слова:** информационная система, управление, процессный подход, инцидент, онтология, классификация, модель предметной области, база знаний, экспертные знания

L. S. Bolotova, A. A. Karasev, V. A. Starykh

## Formalization of the Expert Knowledge for Incident's Management of Information System Based on the Ontological Approach

*This article is devoted to a solution of the problem of collecting and formalization of the expert knowledge representing incidents received during the operation of information systems and designed to address the problem of increasing the efficiency of their management. The existing approaches and tools of a solution of the problem of incident management, their main shortcomings are defined. As perspective approach use of ontology subject domain providing basis for formation of conceptual model of subject domain for incident's management of information system is offered; the organization of objects of subject domain is carried out with use of classification of incidents for information systems with use of structural approach and the qualifier based on the OSI/ISO model. The article also outlines the basic principles of developing a prototype incident management system by using Thinkmap SDK tools.*

**Keywords:** information system, control, process approach, the incident, ontology, classification, the domain model, knowledge base, the expert knowledge

### Введение

Постоянный рост объемов обрабатываемой информации, числа пользователей и повышение требований к производительности и надежности приводят к неизбежному увеличению сложности информационных систем (ИС). Современная ИС является децентрализованным разнородным программно-аппаратным комплексом, требующим для успешного функционирования значительных затрат мате-

риальных, финансовых, временных и человеческих ресурсов. Децентрализация в данном случае означает наличие нескольких территориально распределенных центров хранения и обработки данных (ЦХОД), не всегда связанных друг с другом. Разнородность комплекса объясняется использованием в его составе оборудования и программного обеспечения (ПО) различных производителей, а также наличием разных версий программных продуктов и инстру-

ментов, созданных или доработанных самостоятельно. Ограниченность доступных ресурсов при построении и модернизации также приводит к росту разнородности ИС. Обеспечением ее функционирования при этом занимаются специалисты с разным уровнем квалификации и опытом работы, не всегда обладающие полной информацией о принципах организации работ смежных подсистем.

В этой ситуации особое значение приобретает задача эффективного управления программно-аппаратным комплексом, заключающаяся в управлении процедурами, связанными с обслуживанием аппаратного и программного обеспечения, а также информационными, материальными и временными ресурсами. На сегодняшний день наибольшее распространение получил процессный подход, рассматривающий задачу управления как совокупность взаимосвязанных процессов<sup>1</sup>, выполняющихся на протяжении всего жизненного цикла ИС.

### Типовые решения эффективного управления

Стандартом де-факто процессного подхода к управлению ИС стала библиотека инфраструктуры информационных технологий ITIL (IT Infrastructure Library), первая версия которой была разработана по заказу правительства Великобритании в конце 1980-х годов. Текущая версия библиотеки ITIL v3 2011 Edition [1] объединяет все процессы и функции управления ИС в пять групп: стратегия, проектирование, преобразование, эксплуатация и постоянное улучшение услуг. Особую роль в управлении ИС играют процессы, относящиеся к группе эксплуатации услуг (Service Operation), в первую очередь — процессы управления инцидентами<sup>2</sup> и проблемами<sup>3</sup>.

Алгоритмы выполнения этих процессов, а также программные инструменты, предназначенные для их автоматизации, используются в первую очередь службой технической поддержки, Service Desk в терминологии ITIL. К ее основным функциям относятся регистрация, сбор необходимой информации, классификация и категоризация, разрешение и документирование информации о способах и методах разрешения возникающих инцидентов. В ходе работы служба технической поддержки (СТП) использует как нормативные данные, так и специальные знания об особенностях функционирования

конкретной системы, накопленные в процессе эксплуатации и являющиеся для нее уникальными [3].

Нормативные данные могут быть представлены в виде справочников, спецификаций, инструкций по установке, настройке и эксплуатации серверного и сетевого оборудования, руководств администраторов и пользователей программных продуктов. К их преимуществам относится легкость хранения и передачи между сотрудниками организации; к недостаткам — описание функционирования отдельно взятого объекта, не учитывающее влияние других компонентов существующей инфраструктуры. С этой точки зрения специальные знания, позволяющие учитывать аспекты взаимодействия аппаратных и программных компонентов различных производителей в составе действующей ИС, не всегда отраженные в соответствующей документации разработчика, обладают большей ценностью.

Таким образом, эффективность функционирования ИС (определяемая в том числе числом возникающих инцидентов и временем, затрачиваемым на их обнаружение и разрешение) зависит от возможности накопления (фиксации) и повторного использования информации, полученной в ходе обработки ранее возникавших инцидентов, для сокращения времени разрешения схожих и однотипных инцидентов в дальнейшем. В то же время отсутствие такой информации приводит к возникновению целого ряда факторов, негативно воздействующих на функционирование ИС и возможность управлять ею, в том числе:

- необходимость повторного многократного поиска способа разрешения однотипных инцидентов вместо использования стандартной, ранее задокументированной последовательности действий;
- невозможность заранее устранить причину инцидента на основании имеющейся информации — управляющее воздействие оказывается после возникновения инцидента, т. е. нарушения штатного режима функционирования ИС;
- недостаток управляющей информации, приводящий к принятию решений на основе предположений, а не ранее зафиксированных фактов;
- потеря значительного объема незафиксированных уникальных знаний в случае увольнения сотрудников СТП.

Инструментом, используемым в соответствии с рекомендациями библиотеки ITIL для хранения данных о зарегистрированных в системе инцидентах, причинах их возникновения и способах разрешения, являются *базы данных инцидентов и проблем*. Данные об инцидентах могут поступать из различных источников: от пользователей предоставляемых услуг, сотрудников организации, в автоматическом режиме от установленных систем мониторинга и управления. Обнаруженный любым из перечисленных способов инцидент регистрируется сотрудниками СТП в базе данных инцидентов.

<sup>1</sup> Процесс в данном случае определяется как логически организованная последовательность работ или видов деятельности, направленная на достижение поставленной цели.

<sup>2</sup> Библиотека ITIL определяет термин "инцидент" [2] как любое событие, которое не является частью штатного функционирования ИС и может привести к ухудшению параметров функционирования ИС, отдельной ее подсистемы или службы.

<sup>3</sup> Проблемой [2] называется причина возникновения одного или нескольких инцидентов, не зарегистрированная к моменту их возникновения.

Запись об инциденте содержит набор типовых полей:

- дату и время регистрации;
- источник поступления информации об инциденте;
- первичное описание;
- категорию, определяющую, на какую подсистему ИС или предоставляемую услугу он оказывает влияние;
- приоритет, рассчитываемый на основании степени воздействия и требуемой срочности разрешения, а также ряд других.

Запись в базе данных инцидентов может не содержать описания последовательности действий, необходимых для его разрешения.

Для хранения информации о причинах возникновения и возможных способах разрешения используется база данных проблем. Она содержит не только подробное руководство по разрешению возникающих инцидентов на основе задокументированных ранее в процессе эксплуатации ИС решений, в том числе и обходных, применяющихся для экономии времени и других ресурсов организации, но и рекомендации по устранению первопричин их возникновения.

Проблема, способная одновременно являться как причиной возникновения, так и результатом ранее произошедших в ИС инцидентов, в большинстве случаев связана с более чем одним инцидентом, каждый из которых оказывает сходное влияние на штатное функционирование ИС [4]. Это означает, что для эффективного многократного использования накопленных знаний базы данных инцидентов и проблем должны быть взаимосвязаны — запись об инциденте должна быть связана со всеми записями о проблемах, которые могли привести к его возникновению, а запись о проблеме — со всеми записями о вызванных ею инцидентах. При этом запись о проблеме может также рассматриваться как самостоятельный источник информации, не зависящий от связанных с ней записей об инцидентах. Например, определение первопричин возникновения инцидента — поиск проблемы и документирование известной ошибки — может продолжаться после успешного разрешения самого инцидента и закрытия соответствующей записи.

В процессе эксплуатации ИС периодически возникают однотипные инциденты, действия для разрешения которых хорошо известны. При регистрации других используется стандартная процедура привязки классификационных данных инцидента к известным ошибкам и записям о проблемах. В случае обнаружения в базе данных записи об инциденте с аналогичным описанием выполняется последовательность ранее зафиксированных и строго регламентированных действий, что позволяет сократить время и значительно упростить процесс его разрешения. Классификация инцидентов и проблем

позволяет сотрудникам службы технической поддержки рассматривать базы данных проблем как набор информации, организованный и хранимый таким образом, чтобы к нему можно было оперативно обращаться с помощью ссылок по легко определяемым типичным признакам новых инцидентов.

Во многих программных продуктах, предназначенных для автоматизации процессов управления инцидентами и проблемами, базы данных объединяются в общую постоянно пополняемую базу знаний ИС организации. Такая база содержит информацию о зарегистрированных инцидентах, определенных способах их разрешения, в том числе обходных решениях, установленных причинах возникновения, стандартной последовательности действий для их устранения, а также все существующие связи между ними и объектами управления. Подобные базы знаний могут быть как централизованными, предназначенными для использования в рамках одной организации или организационного подразделения, так и распределенными, предоставляющими интерфейсы для интеграции в открытые репозитории знаний в области управления ИС.

Для определения основных подходов к накоплению и последующему использованию специальных знаний в процессе функционирования ИС был проведен сравнительный анализ ряда специализированных программных продуктов<sup>4</sup>. Анализ проводился по таким критериям, как источники наполнения баз данных, наличие предустановленных баз знаний, возможности создания и поддержания в актуальном состоянии связей между записями об инцидентах, проблемах и объектах управления, а также наличие встроенных классификаторов. В результате (фрагмент сравнительного анализа для трех программных продуктов приведен в таблице) был выявлен ряд недостатков существующих программных продуктов и применяемых в них методов, в том числе:

- недостаточная формализация. Собранные знания хранятся в виде самостоятельных, не связанных друг с другом информационных статей, содержащих слабоструктурированную информацию, что не позволяет эффективно использовать информационные технологии для ее поиска и обработки;
- отсутствие встроенных средств классификации. В большинстве программных продуктов отсутствуют классификаторы, в соответствии с которыми могли бы распределяться новые записи об инцидентах. В результате этого усложняется первоначальная привязка инцидентов и множества затронутых объектов управления, а также определение причин их возникновения, повышается вероятность появления дублирующихся записей и записей с однотипным содержанием;

<sup>4</sup> Рассматривалось ПО автоматизации процессов управления инцидентами и проблемами IBM Tivoli Monitoring, Microsoft System Center Service Manager, HP Service Manager, BMC Remedy IT Service Management Suite, OTRS ITSM.

**Сравнительный анализ программных продуктов, применяемых для фиксации специальных знаний в процессе эксплуатации ИС**

Название системы	IBM Tivoli Monitoring	Microsoft System Center Service Manager [5]	HP Service Manager Knowledge Management [6]
Формат хранения данных в базе инцидентов	Файлы .htm	СУБД	СУБД
Источники наполнения базы инцидентов	Поставляется с предустановленным набором типовых инцидентов, может пополняться вручную администратором	Системы мониторинга, приложения сторонних производителей, e-mail обращения пользователей может пополняться вручную администратором	Может пополняться вручную администратором или пользователями при наличии необходимых прав
Наличие предустановленного набора типовых инцидентов	Да	Набор шаблонов для создания записей об инцидентах	Нет
Возможность редактирования связей между инцидентами	Не предусмотрена	Не предусмотрена	Не предусмотрена
Графическое отображение инцидентов и связей	Нет	Нет	Нет
Поиск в базе инцидентов	Нет	Полнотекстовый поиск	Полнотекстовый поиск по всем документам, преобразованным в html
Встроенные классификаторы инцидентов	Отсутствуют	Отсутствуют	Отсутствуют

- отсутствие связей. В рассмотренном ПО не реализованы инструменты создания, редактирования и поддержания в актуальном состоянии связей между инцидентами, проблемами и объектами управления. Это приводит к отсутствию у пользователя возможности видеть причинно-следственные связи между событиями, приводящими к нарушению штатного режима функционирования ИС.

Как было отмечено выше, повторное использование накопленных специальных знаний позволяет повысить эффективность управления ИС за счет сокращения времени разрешения возникающих инцидентов и принятия управленческих решений, повышения точности этих решений, что в целом позволяет оптимизировать использование временных, финансовых, людских и других ресурсов организации. Однако такое повышение возможно только в случае решения задач формализации и классификации специальных знаний в процессе эксплуатации ИС.

Перспективным подходом к их решению является использование онтологий. Онтология — термин, в интеллектуальных информационных системах и технологиях инженерии знаний обозначающий систему понятий заданной предметной области, представляемой в виде набора значимых сущностей и отношений, возможных между ними. Онтологии используются для формального определения понятий и свойств, формирующих концептуальную модель предметной области.

Появление и стремительное развитие онтологий на сегодняшний день связано с необходимостью перехода на качественно новый уровень поиска и обработки информации при решении таких задач, как:

- обработка значительных объемов слабоструктурированной, часто противоречивой информации в различных областях деятельности;

- совместное использование общего хранилища информации специалистами в различных областях, в том числе смежных, для достижения различных целей;
- структуризация совместно используемой информации для представления пользователям;
- сокращение времени поиска требуемой информации и повышение точности, соответствия результатов поиска поисковому запросу.

Существующие определения термина "онтология" значительно различаются и зависят от области применения. Так, с точки зрения информационных систем онтология является базой знаний специального типа, которая может пополняться, использоваться, отчуждаться от разработчика, использоваться совместно с другими базами знаний или разделяться их пользователями [7].

Наиболее распространенное определение, используемое в интеллектуальных системах, описывает онтологию как формальную точную спецификацию совместно используемой концептуализации [8]. В этом определении под концептуализацией подразумевается абстрактная модель определенной предметной области, для которой установлены понятия (концепты), необходимые для описания сущностей, явлений и их взаимосвязей. Формальность означает возможность использования информационных технологий и средств вычислительной техники для автоматизации обработки данных онтологии. Точность подразумевает использование явно определенных типов включенных в состав онтологии понятий и ограничений на область их применения. Совместное использование означает, что онтологии используются для описания общепринятых в определенной группе или сообществе специалистов знаний на основе словаря терминов с объемом, достаточным для их описания.

Для всех определений термина "онтология" общим является наличие словаря имен сущностей, используемого для обмена знаниями в определенной предметной области, и множества связей между ними [9]. Связи могут относиться как к универсальным типам — "часть — целое", "причина — следствие", так и к типам, специфичным для данной предметной области. Сущности, как и связи, могут обладать заданными свойствами, необходимыми для отражения в абстрактной модели значимых свойств и отношений объектов области знаний.

Анализ возможностей онтологического подхода к организации знаний [8, 10] позволяет сделать вывод о его применимости к формализации и структурированию экспертных знаний предметной области управления инцидентами, накопленными в ходе эксплуатации ИС.

Следующим шагом в решении поставленных задач является выбор принципов классификации накопленных знаний, в соответствии с которыми они будут организованы в онтологии. В данном исследовании применяются два независимых классификатора объектов предметной области управления ИС. Первый основан на структурном подходе теории информационных систем, позволяющем учитывать универсальные иерархические связи между объектами. Второй использует эталонную модель взаимодействия открытых систем (open systems interconnection, OSI) для определения специфичных для предметной области связей между объектами и их свойств.

Структурный подход базируется на утверждении о том, что любая ИС может рассматриваться как совокупность своих подсистем. Примерами таких подсистем являются структурированная кабельная система (СКС), серверное оборудование, виртуальная инфраструктура или почтовая подсистема организации. ИС может состоять из произвольного числа подсистем; одна и та же подсистема может входить в состав ИС многих организаций. Далее в каждой подсистеме выделяется множество программно-аппаратных компонентов, из которых она состоит. Такими компонентами являются физическое или виртуальное оборудование и программное обеспечение. Следующим этапом декомпозиции является выделение в составе каждого аппаратно-программного компонента объектов управления, являющихся атомарными с точки зрения установления взаимосвязей с инцидентами предметной области (в данном случае понятие объектов управления аналогично понятию конфигурационных

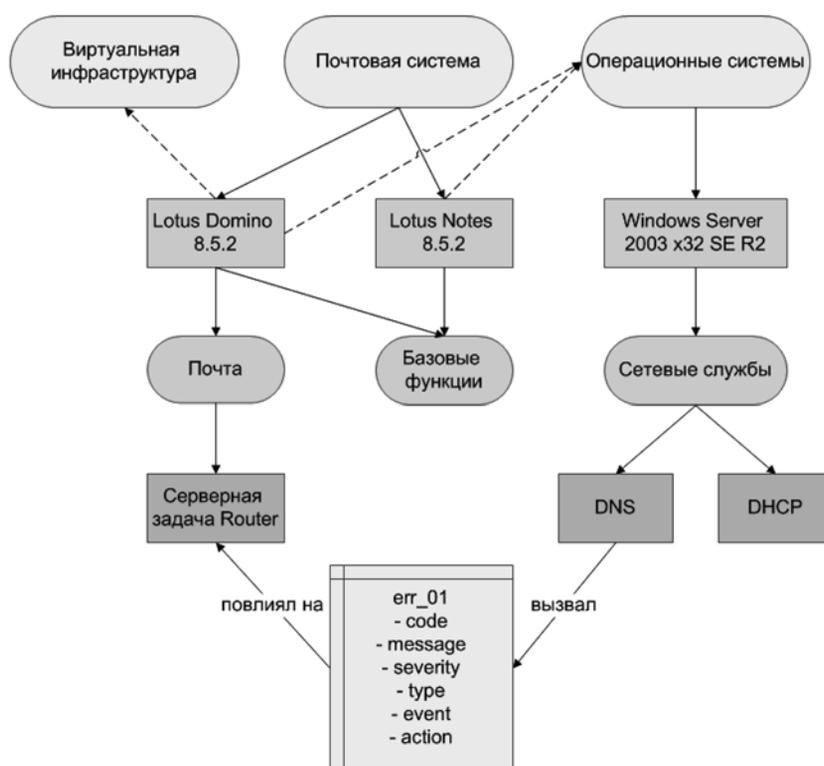


Рис. 1. Связи инцидента с объектами управления при использовании структурного подхода

единиц (configuration items), используемых процессом управления конфигурациями библиотеки ИТІЛ. Атомарный объект либо является причиной возникновения инцидентов, либо оказывается затронутым инцидентами, вызванными другими объектами. При необходимости объекты управления в составе аппаратно-программного комплекса могут объединяться в группы по тем или иным признакам.

На рис. 1 представлен пример классификации инцидента в ИС с использованием структурного подхода.

Инцидент, вызванный сбоем в работе сетевой службы DNS операционной системы Windows Server 2003 и влияющий на доставку почтовых сообщений серверной задачей Router ПО Lotus Domino связан с соответствующими атомарными объектами. Свойства объектов предметной области и отношения между ними, описываемые в рамках онтологического подхода, могут использоваться для фиксации знаний как о самом инциденте, так и обо всех инцидентах, с ним связанных.

Эталонная модель OSI предназначена для концептуального определения процесса взаимодействия открытых систем при последовательном рассмотрении его на семи уровнях, начиная с физического и заканчивая уровнем приложений. Назначение каждого уровня стандартизируется, определяется набор протоколов, функционирующих в его рамках, описываются интерфейсы, позволяющие передавать данные от одного уровня к другому.

Классификатор, основанный на эталонной модели взаимодействия открытых систем, позволяет логически упорядочить, сгруппировать и распределить инциденты по ее уровням. Распределение инцидентов может выполняться в зависимости как от того, к какому уровню модели относится объект управления, связанный с ними<sup>5</sup>, так и от того, на каком уровне было нарушено взаимодействие в результате возникновения инцидента. В качестве примера распределения инцидентов по уровням модели OSI можно привести:

- физический уровень — нарушение целостности оптоволоконного кабеля, соединяющего центры обработки данных;
- канальный уровень — сбой сетевого подключения, вызванный неисправностью сетевого интерфейса сервера;
- сетевой уровень — не найден путь к указанному серверу из-за неверной конфигурации шлюза по умолчанию;
- транспортный уровень — сбой обмена TCP-пакетами, вызванный некорректной настройкой параметра MTU (Maximum Transmission Unit, максимальный размер полезного блока данных);
- сеансовый уровень — невозможность установления соединения с удаленным сервером, вызванная использованием различных версий протокола SSH (Secure Shell);
- представительский уровень — ошибка подключения к удаленному рабочему столу при использовании клиентом и сервером несовместимых уровней шифрования;
- прикладной уровень — запрашиваемая пользователем Web-страница недоступна из-за ошибки сервера.

Использование классификатора, основанного на семиуровневой модели взаимодействия OSI, позволяет организовывать и соответствующим образом распределять накопленные знания о причинах возникновения и способах разрешения, а также о существующих взаимосвязях между инцидентами. Это дает специалисту, работающему с онтологией, возможность последовательно просматривать уровни, начиная с физического и заканчивая уровнем приложений в поисках корневой причины возникновения инцидента (рис. 2).

В качестве инструмента построения и графического отображения инцидентов и связей между ними в данной работе используется комплект средств разработки Thinkmap [11]. Он обеспечивает поддержку ряда современных технологий (Java EE, HTML5, CSS, XML) и предназначен для автоматизации обработки и визуализации (в том числе построения графических интерфейсов приложений)

<sup>5</sup> Штатное функционирование которого как объекта управления было нарушено инцидентом либо само привело к возникновению инцидента.



Рис. 2. Распределение причин возникновения инцидента по уровням модели OSI

как формализованных, так и описываемых с помощью естественных языков данных. Модели предметной области, построенные с использованием Thinkmap, учитывают связи и различные типы взаимодействия объектов, а также их свойства. В качестве источника данных используются XML файлы, файлы данных (flat file) или реляционные СУБД; данные могут располагаться на выделенном сервере или локально на компьютере пользователя.

### Определение структур данных

Использование классификатора, основанного на выделении подсистем в составе ИС, требует определения структур данных, предназначенных для описания каждого уровня вплоть до атомарных объектов управления. Примером такой структуры для аппаратно-программного компонента в формате XML является:

```
<nodetype name = "COMPONENT" >
  <property name = "compname" type = "String" />
  <property name = "vendor" type = "String" />
  <property name = "comment" type = "String" />
</nodetype>
```

Свойство "compname" используется для хранения названия серии и модели оборудования или названия и версии ПО; "vendor" — компании-производителя. Свойство "comment" содержит подробную характеристику, позволяющую специалисту, не знакомому с тем или иным компонентом ИС,

быстро определить его тип и принадлежность к определенной подсистеме.

Далее приведен пример записи о коммутаторе Cisco как аппаратном компоненте подсистемы "активное сетевое оборудование":

```
<node type = "COMPONENT" id = "121" >
  <property name = "compname" value = "Catalyst 3750G" />
  <property name = "vendor" value = "Cisco" />
  <property name = "comment" value = "Коммутатор" />
</node>
```

Основным типом сущностей в рассматриваемой онтологии предметной области управления ИС является инцидент. Приведенные выше классификаторы используются в первую очередь для организации и привязки к соответствующим объектам управления (узлам классификаторов) записей об инцидентах. Структура инцидента описывается в данной работе следующим образом [12]:

```
<nodetype name = "INCIDENT" >
  <property name = "code" type = "String" />
  <property name = "message" type = "String" />
  <property name = "severity" type = "String" />
  <property name = "type" type = "String" />
  <property name = "event" type = "String" />
  <property name = "action" type = "String" />
  <property name = "osi" type = "Float" />
  <property name = "comment" type = "String" />
</nodetype>
```

Свойство "code" здесь предназначено для хранения идентификатора инцидента, отображаемого в графическом интерфейсе пользователя, построенного с использованием средств разработки Thinkmap. Свойство "message" содержит стандартное сообщение об ошибке, генерируемое соответствующим объектом управления. Поле "severity" определяет приоритет инцидента, вычисляемый на основании степени его воздействия на ИС организации и требуемой срочности разрешения. Поле "type" позволяет определить, с какой подсистемой, группой решаемых задач или аппаратно-программным компонентом связан инцидент. Свойства "event" и "action" содержат, соответственно, подробное описание самого инцидента и руководство по его разрешению, представляющее собой рекомендуемую последовательность действий, выполненных и зафиксированных при первоначальном разрешении инцидента. Поле "osi" является служебным и определяет принадлежность инцидента к заданному уровню в соответствии с классификатором, основанным на эталонной модели взаимодействия открытых систем. Свойство "message" содержит краткое описание инцидента, также доступное в графическом интерфейсе.

Далее приведен пример записи об инциденте в соответствии с рассмотренной структурой в формате XML:

```
<node type = "INCIDENT" id = "171021" >
  <property name = "code" value = "INC_0038" />
  <property name = "message" value = "The remote procedure call timed out and was canceled" />
  <property name = "severity" value = "Средняя" />
  <property name = "type" value = "Администрирование ОС" />
  <property name = "event" value = "При попытке выполнения удаленного вызова процедуры Active Directory (репликация изменений) время ожидания истекло, вызов отменен" />
  <property name = "action" value = "Установите значение параметра реестра RPC Replication Timeout (HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\NTDS\Parameters), равное 45 минутам" />
  <property name = "osi" value = "7" />
  <property name = "comment" value = "Ошибка репликации изменений службы каталогов Active Directory" />
</node>
```

Для описания связи атомарного объекта управления с одним или несколькими инцидентами используется следующая структура данных:

```
<edgetype name = "OBJECT_INCIDENT" from = "OBJECT" to = "INCIDENT" >
  <property name = "issuedby" type = "Boolean" />
  <property name = "affect" type = "Boolean" />
  <property name = "comment" type = "String" />
</edgetype>
```

Здесь поля "issuedby" и "affect" определяют, стал ли объект управления причиной возникновения инцидента или, в свою очередь, было ли его функционирование нарушено возникшим инцидентом. Свойство "comment" содержит текстовое пояснение, доступное в графическом интерфейсе.

### **Работа прототипа системы управления инцидентами**

На рис. 3 представлен общий вид пользовательского интерфейса прототипа системы управления инцидентами, разработанной с использованием инструмента Thinkmap.

В рабочей области расположены инциденты, обозначенные идентификаторами, хранящимися в поле "code" соответствующей записи, связи между ними и объектами управления и сами объекты управления, отображенные в зависимости от текущего представления. Для выбранного инцидента в правой панели отображается подробная информация в со-



Г. Н. Лебедев<sup>1</sup>, д-р техн. наук, проф., e-mail: k301@mail.ru, М. Е. Семенов<sup>2</sup>, д-р физ.-мат. наук, проф.,  
Д. В. Грачиков<sup>3</sup>, ассистент кафедры, О. И. Канищева<sup>2</sup>, канд. физ.-мат. наук, доц.

<sup>1</sup> Московский авиационный институт

<sup>2</sup> Военный учебно-научный центр Военно-воздушных сил Военно-воздушной академии  
им. проф. Н. Е. Жуковского и Ю. А. Гагарина

<sup>3</sup> Воронежский государственный университет

## Гистерезисная модель синхронизации биологических нейронов\*

*Приводится модель синхронизации биологических нейронов с учетом гистерезисных свойств метаболитных рецептивных кластеров. Изучаются свойства предложенной модели, рассматриваются различные варианты связи между отдельными нейронами.*

**Ключевые слова:** нейронные сети, гистерезис, моделирование, аттрактор, дифференциальные уравнения с запаздыванием

G. N. Lebedev, M. E. Semenov, D. V. Grachikov, O. I. Kanishcheva

## Hysteretic Synchronization Models of Biological Neurons

*The article provides a synchronization model of biological neurons based on hysteretic properties of metabotropic receptive clusters. We study the properties of the proposed model, various variants connections between individual neurons.*

**Keywords:** neural networks, hysteresis, modeling, attractor, differential equations with delay

### Введение

В настоящее время в теории нейронных сетей активно развивается направление по изучению осцилляторных аспектов функционирования мозга. Существует ряд моделей [1–6], которые в силу разной степени своей биологической обоснованности способны описывать взаимодействие нейронов в коре головного мозга. Особо отметим наиболее приближенную к биологическим данным модель Ходжкина—Хаксли [4, 5], которая из-за своей сложности практически не применяется для моделирования нейронных сетей. Центральное место в моделях отводится задачам, способным дать ответы на важные вопросы психологии и нейробиологии. Учитывая, что к настоящему времени вычислительные мощности неуклонно растут, для исследования нейронных сетей можно использовать модели, максимально приближенные к биологическим данным, так как человеческий мозг в решении множества задач остается эталоном.

В работе проводится исследование поведения биологических нейронных сетей, основанных на модели Кашенко — Майорова [1]. Однако, в отличие от этой модели, в настоящей работе организация связи между отдельными нейронами модели-

руется с учетом биологических особенностей, подробно рассмотренных в работе А. Н. Радченко [7]. Согласно современным представлениям нейрон окружен клеточными образованиями (кластерами), которые при определенных условиях способны запустить эндогенные (внутренние) процессы в нейроне, за которыми может последовать спайк. Запуск эндогенных процессов имеет гистерезисную природу. Для моделирования межнейронных связей используется операторная трактовка гистерезисных нелинейностей, которая была впервые введена М. А. Красносельским и А. В. Покровским в работе [8], позднее этот подход был развит применительно к системам, содержащим звенья гистерезисной природы в работе [9]. Предложенная модель была выбрана в силу ее биологической обоснованности. Также выявляются закономерности в динамике нейронов в зависимости от структуры нейронных сетей и связей между нейронами.

### Базовая модель

Феноменологическая модель Кашенко — Майорова основана на течении через мембрану нейронов калиевых и натриевых токов. Она была успешно использована для моделирования кольцевых нейронных структур, в которых возбуждение нейронов-предшественников волнообразно передавалось на нейроны-последователи [1]. Важным свойством этой модели является то, что формализуется био-

\* Работа выполнена при поддержке РФФИ, гранты № 12-07-00252-а, 13-08-00532-а.

логический факт запаздывания калиевых токов от натриевых.

В этой модели функция активации нейрона описывается дифференциальным уравнением с запаздыванием:

$$\dot{u} = \lambda(-1 + fK(u(t-1))) - fNa(u)u, \quad (1)$$

с соответствующим начальным условием:

$$u(t)|_{-1 \leq t \leq 0} = \varphi(t), \quad \varphi(t) \leq Ce^{-1-\varepsilon},$$

где  $\lambda$  — коэффициент,  $C$  — константа, обусловленные модельными представлениями о биологических свойствах мембраны нейрона, функции  $fK(u)$  и  $fNa(u)$  характеризуют калиевый и натриевый токи. На них, как правило, накладываются ограничения:

$$\begin{aligned} fK(u) > 0, fNa(u) > 0, \\ -1 - fNa(0) + fK(0) > 0, \\ fK(u) < C\lambda e^{-1-\varepsilon}, \\ fNa(u) < C\lambda e^{-1-\varepsilon}, \end{aligned} \quad (2)$$

где  $\varepsilon > 0$ , параметр, обусловленный модельными представлениями.

Нейрон может воспринимать как электрическую, так и химическую стимуляцию. В случае электрического возбуждения функция активации нейрона будет удовлетворять уравнению

$$\dot{u} = \lambda(-1 + fK(u(t-1))) - fNa(u)u + g(t), \quad (3)$$

где  $g(t)$  — интенсивность электрического воздействия.

В случае химического воздействия

$$\dot{u} = \lambda(-1 + fK(u(t-1))) - fNa(u)u + v(t)u, \quad (4)$$

где  $v(t)$  — интенсивность химического воздействия.

Согласно [1], электрическое воздействие эффективно навязывает период генерации спайков нейрона, в то время как химическое эффективнее при моделировании связи и синхронизации ансамблей в нейронной сети.

### Модель памяти нейронов Радченко

Способность нейронной сети воспринимать и хранить информацию связана с обучением. В классических искусственных нейронных сетях эта задача решается подбором и корректировкой весов для связей между нейронами по определенным закономерностям. В работе [7] была предложена новая модель, в которой структура связей между нейронами основана на биологических данных и имеет гистерезисную природу. Там же приводятся результаты биологических экспериментов, в ходе которых обнаружены специальные образования вокруг синапсов нейронов (место, в котором нейрон воспринимает воздействие) — метаботропный рецептивный кластер (МРК). Это образование при воздействии на нейрон может запустить эндогенные (внутренние) химические процессы, тем самым, спровоцировав спайк (выброс энергии, всплеск мембранного потенциала). Гистерезисная природа запуска эндогенных процессов зависит от вида воздействия на нейрон (химическое или электрическое).

**Химическое воздействие на МРК.** При химическом воздействии на нейрон запуск внутренних процессов посредством МРК будет происходить согласно рис. 1.

Изображенные на рис. 1 кривые описываются уравнением:

$$u = By \sqrt{\frac{1-k}{y-k}} - Ay^2, \quad (5)$$

где  $y$  — относительное расстояние между зарядами в МРК;  $A = \frac{\lambda_0 \rho}{\varepsilon}$ ,  $B = \lambda_0 \sqrt{\frac{2Y}{\varepsilon}}$ ;  $\rho$  — концентрация зарядов в МРК. Экспериментально вычисленные значения констант определяются равенствами:

$$\lambda_0 = 1,4 \cdot 10^{-9}, \quad \varepsilon = 8,85 \cdot 10^{-11}, \quad Y = 0,67 \cdot 10^6.$$

Форма кривой (рис. 1) определяется коэффициентом  $0 \leq k \leq 1$ . С биологической точки зрения этот параметр характеризует подвижность зарядов в МРК.

При росте воздействия  $u_m$  на МРК происходит переход с верхней части кривой на нижнюю (рис. 1). С биологической точки зрения произойдет химический конформационный переход (ХКП), т. е. частицы рецептивного кластера уплотняются, тем самым аккумулируя энергию. При последующем росте  $u_m$  коэффициент  $k$  приближается к единице и, как следствие, уплотнение МРК продолжится, при этом кривая будет меняться от  $A_1$  до  $A_3$ .

При уменьшении  $u_m$  до значения, достаточного для обратного перехода с нижней части кривой на верхнюю произойдет деполяризационный конформационный переход (ДКП), при этом на нейрон будет направлена увеличенная в сотни раз аккумулированная энергия в МРК после ХКП.

При описанном механизме работы МРК ясно, что наиболее эффективным является периодическое химическое воздействие с растущей амплитудой, подобное воздействию со стороны нейронного ансамбля, который притягивает к себе нейроны, тем самым увеличивая силу притяжения.

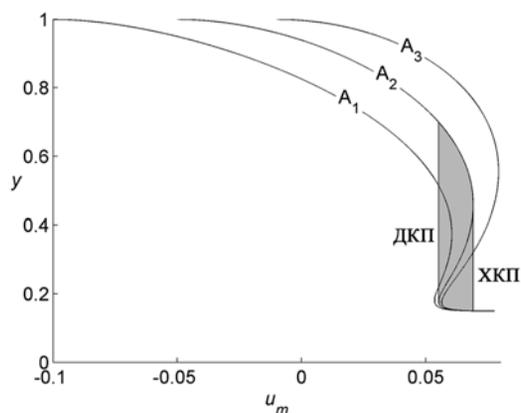


Рис. 1. Изменение формы гистерезисных кривых, характеризующих динамику МРК при химическом воздействии

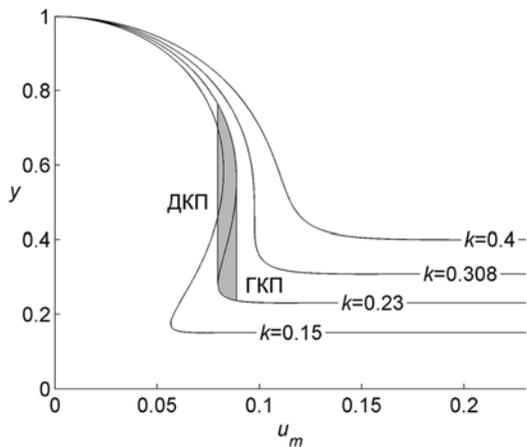


Рис. 2. Изменение формы гистерезисных кривых, характеризующих динамику МРК при электрическом воздействии

**Электрическое воздействие.** Если нейрон подвергается электрическому возбуждению, то уравнение (5) трансформируется и принимает вид

$$u = By \sqrt{\frac{1-k}{y-k}}. \quad (6)$$

Изменение формы кривых в зависимости от коэффициента  $k$  показано на рис. 2 [7].

При увеличении силы электрического воздействия на нейрон уменьшается относительное расстояние между зарядами в МРК и происходит гиперполяризационный переход (ГКП). При уменьшении же силы воздействия, аналогично случаю с химическим внешним воздействием, произойдет ДКП.

Как видно из рис. 2, при уменьшении подвижности зарядов в МРК, чему соответствует увеличение  $k$ , МРК теряет свою гистерезисную природу и утрачивает способность к запуску эндогенных процессов в нейроне. Уменьшение подвижности зарядов может быть вызвано длительным увеличением  $u_m$ . Таким образом, сильная электрическая стимуляция ослабляет воздействие МРК на нейрон.

### Описание модели

Ниже рассматривается нейронная сеть, отдельные элементы которой описываются дифференциальными уравнениями:

$$\dot{u}_i = \lambda(-1 + fK(u_i(t-1)) - fNa(u_i) + Y_i) + I_i, \quad (7)$$

где  $u_i$  — мембранные потенциалы нейронов;  $Y_i$  — воздействие, оказываемое на  $i$ -й нейрон со стороны нейронной сети;  $I_i$  — внешнее воздействие на  $i$ -й нейрон. Здесь  $Y_i$  отвечает за химическую связь элементов сети, а  $I_i$ , в свою очередь, моделирует электрическое внешнее воздействие на нейроны. Положим

$$Y_i = \begin{cases} y_i, & \text{если } \dot{y}_i > \gamma; \\ 0, & \text{в остальных случаях,} \end{cases} \quad (8)$$

где  $\gamma$  — пороговая скорость роста  $y_i$ . Здесь ограничение введено для определения этапа ХКП, когда  $y_i$  быстро растет. Связь между нейронами количественно будем характеризовать соотношением

$$x_i(t) = \sum_{j=1}^N D_{ij} \frac{\int_{t-T}^t |u_i - u_j| ds}{\int_{t-T}^t u_i ds}, \quad (9)$$

где  $D_{ij}$  — коэффициент силы связи между  $i$ -м и  $j$ -м нейронами;  $N$  — общее число нейронов. Параметр  $T$  подбирается таким образом, чтобы на соответствующем промежутке содержался ровно один спайк. По определению, значение величины  $x_i(t)$  возрастает вместе с интегральной мерой рассинхронизации нейронов.

В свою очередь,  $x_i(t)$  является входом гистерезисного преобразователя  $\Gamma[x_0, y_0]$ , т. е. связь между  $y_i$  и  $x_i$  определим следующим образом:

$$\dot{y}_i = f(x_i, y_i); \quad y_i(0) = y_0, \quad (10)$$

где функция  $f(x_i, y_i)$  задается соотношением

$$f(x_i, y_i) = \chi \left( \operatorname{arctg} \left( \frac{\xi(y_i - \rho)^3 + 4(1 - k_i)(y_i - \rho)}{2} \right) + \frac{\pi}{2} \right) - x_i. \quad (11)$$

Здесь  $\chi = 50$ ,  $\xi = -4$ ,  $\rho = 0,6$  — константы, характеризующие форму гистерезисной петли. Линия уровня  $f(x_i, y_i) = 0$  приведена на рис. 3.

Электрическое воздействие от МРК определим следующим образом:

$$I_i = \begin{cases} y_i, & \text{если } \dot{y}_i > \gamma; \\ 0, & \text{в остальных случаях;} \end{cases} \quad (12)$$

$$\dot{y}_i = f(x_i, y_i), \quad y_i(0) = y_0; \quad (13)$$

$$x_i = Ae^{-\alpha t} \sin(t - g_i), \quad (14)$$

где  $g_i$  — внешнее воздействие на  $i$ -й нейрон;  $A$  — положительная константа;  $\alpha$  — коэффициент, определяющий скорость потери гистерезисных свойств МРК.

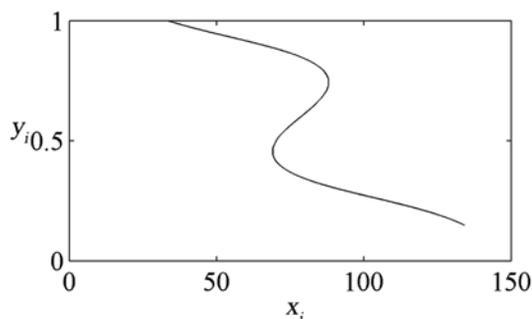


Рис. 3. Линия уровня  $f(x_i, y_i) = 0$

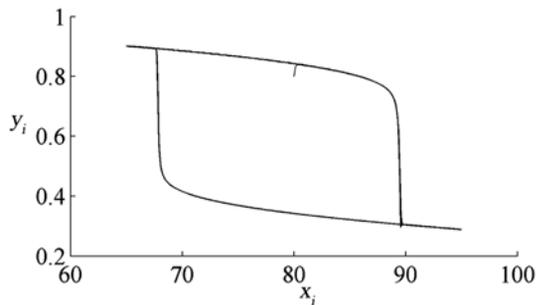


Рис. 4. Динамика  $y_i$  при периодическом входе

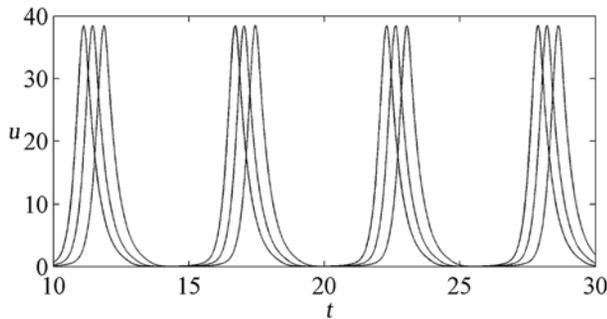


Рис. 5. Динамика мембранных потенциалов трех нейронов с коэффициентами связи  $D_{ij} = 0,01$

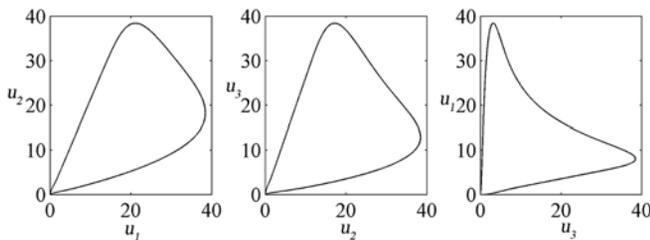


Рис. 6. Проекция фазового портрета системы (7) при  $D_{ij} = 0,01$

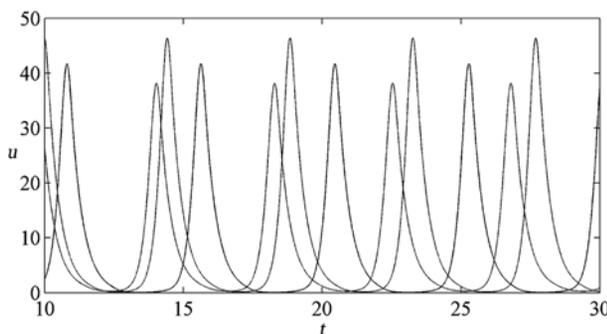


Рис. 7. Динамика трех нейронов при выполненном условии (16)

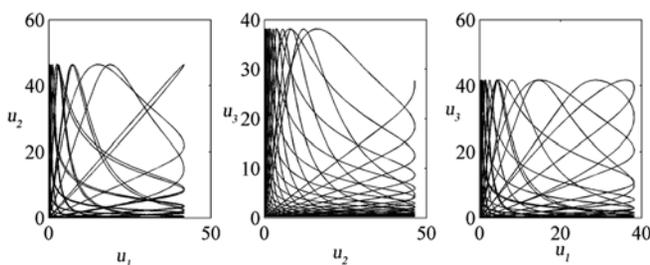


Рис. 8. Фазовые портреты (6) при выполненном условии (16)

Гистерезисная петля преобразователя (10) с периодическим входом  $x_i = 15\sin(t) + 80$  приведена на рис. 4.

Как следует из результатов работы [7], МРК имеет способность инициировать спайк нейрона во время ДКП. То есть пока воздействие на нейрон растет или остается постоянным, нейрон будет реагировать на него без посредников. Однако в случае уменьшения этого воздействия спайк нейрона будет обусловлен МРК.

## Результаты

**Исследование полностью связанной нейронной сети со "слабой" связью.** Рассмотрим нейронную сеть, состоящую из трех нейронов с входными воздействиями (12) на каждый ее элемент, при этом полагаем, что внешние воздействия удовлетворяют условию

$$0 \leq g_i \leq 1. \quad (15)$$

В приводимом ниже примере эти параметры полагались равными 0,1, 0,5, 0,8 соответственно.

Сначала рассмотрим случай, когда коэффициенты  $D_{ij}$  в соотношении (9) достаточно малы, что нивелирует влияние связи между отдельными нейронами на динамику сети в целом. Результаты моделирования представлены на рис. 5.

Проекция фазового портрета на соответствующие плоскости приведены на рис. 6.

Как видно из рисунков, эффект синхронизации отсутствует на относительно продолжительном временном промежутке.

При отсутствии гистерезиса во входном воздействии, когда внешнее воздействие определялось соотношением

$$I_i = g_i, \quad (16)$$

динамика сети будет иметь вид, иллюстрируемый рисунками 7, 8.

Из сравнения графиков 5, 6 и 7, 8 видно, что гистерезис во внешнем воздействии вносит упорядочивающий эффект в нейронную сеть, однако синхронизация наблюдается лишь для достаточно больших промежутков моделирования.

**Исследование полностью связанной сети с "сильной" постоянной гистерезисной связью.** Рассмотрим сеть из трех нейронов, описываемую системой дифференциальных уравнений (7). Положим, что гистерезисная связь между элементами сети активна и неизменна, т. е.

$$k_i = \text{const}, \quad 0 < k_i < 1.$$

Приведем результаты численного моделирования описанной нейронной сети.

Как видно из рис. 9, при постоянной гистерезисной связи один из нейронов синхронизируется с другим, начиная с 10-й секунды, полная синхронизация начинается примерно с 90-й секунды. Этот

## Заклучение

Рассмотренная в работе модель нейронной сети показывает, что даже при небольшом числе нейронов ее динамика согласуется с биологическими данными [12]. Весомым преимуществом предложенной модели является ее "упорядоченная" реакция на внешние воздействия. Эта особенность, обусловленная учетом гистерезисной связи между отдельными нейронами, позволяет моделировать синхронизацию отдельных нейронов в ансамбли на основе биологических законов [14, 15]. Известно, что модели импульсных нейронных сетей эффективно применяются к решению прикладных задач [16], поэтому можно полагать, что описанная в работе модель биологической нейронной сети может найти применение в задачах сегментации и классификации с эффективностью, превышающей стандартные численные алгоритмы работы с данными.

## Список литературы

1. **Кашенко С. А., Майоров В. В.** Модели волновой памяти. М.: УРСС, 2009. 288 с.
2. **FitzHugh R.** Impulses and physiological states in theoretical models of nerve membrane // Biophysical journal. 1961. V. 1, N 6. P. 445–466.
3. **Nagumo J., Arimoto S., Yoshizawa S.** An active pulse transmission line simulating nerve axon // Proceedings of the IRE. 1962. V. 50, N 10. P. 2061–2070.
4. **Hodgkin A. L., Huxley A. F.** The dual effect of membrane potential on sodium conductance in the giant axon of *Loligo* // The Journal of physiology. 1952. V. 116, N 4. P. 497–506.
5. **Hansel D., Mato G., Meunier C.** Phase dynamics for weakly coupled Hodgkin-Huxley neurons // EPL (Europhysics Letters). 1993. V. 23, N 5. P. 367.
6. **Yu G., Slotine J. J.** Visual grouping by neural oscillators // arXiv preprint arXiv:0807.2928. 2008.
7. **Радченко А. Н.** Ионотропные и метаболитные реакции нейрона как инструменты нейронной памяти // Нейроинформатика. 2006. Т. 1. С. 197–227.
8. **Красносельский М. А., Покровский А. В.** Системы с гистерезисом. М.: Наука, 1983. 271 с.
9. **Семенов М. Е., Матвеев М. Г., Шевлякова Д. В., Канищева О. И.** Зоны устойчивости и периодические решения перевернутого маятника с гистерезисным управлением // Мехатроника, автоматизация, управление. 2012. № 11. С. 8–14.
10. **Казанович Я. Б., Борисюк Р. М.** Синхронизация в нейронной сети фазовых осцилляторов с центральным элементом // Математическое моделирование. 1994. № 6–8. С. 45–60.
11. **Феофан И., Крюков В. И.** Модель внимания и памяти, основанная на принципе доминанты и компараторной функции гиппокампа // Журнал высшей нервной деятельности им. И. П. Павлова. 2004. Т. 54, № 1. С. 11–31.
12. **Семенов М. Е., Грачиков Д. В., Канищева О. И.** Синхронизация нейронных ансамблей под действием МРК // XIII Междунар. конф. "Информатика: проблемы, методология, технологии", 2013.
13. **Абарбанель Г. Д., Рабинович М. И., Селверстон А. А.** и др. Синхронизация в нейронных ансамблях // Успехи физических наук. 1996. Т. 166, № 4. С. 363–390.
14. **Еременко Ю. И., Полещенко Д. А., Глушенко А. И.** Об особенностях практической реализации схемы пид-нейрорегулятора с самонастройкой для управления печами нагрева // Приборы и системы. Управление. Контроль. Диагностика. 2012. № 1. С. 25–30.
15. **Гулаков В. К., Огурцов С. Н., Трубаков А. О.** Сегментация пейзажных изображений // Информационные технологии. 2013. № 1. С. 40–45.
16. **Мельников И. И., Демиденков К. А., Емельянов И. А., Евсеев И. А.** Детектор движения на основе импульсных нейронных сетей // Информационные технологии, 2013. № 7. С. 57–60.

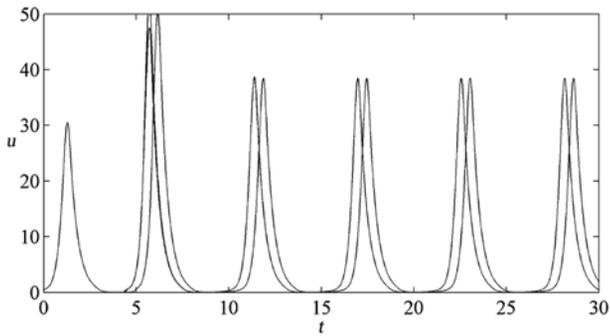


Рис. 9. Динамика нейронной сети с входными воздействиями 0, 0,5, 1

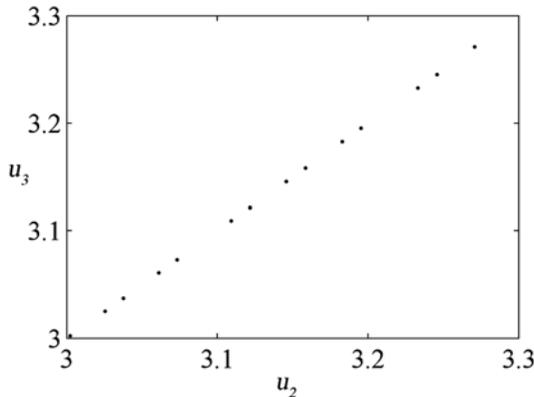


Рис. 10. Сечение Пуанкаре для системы из трех нейронов с входными воздействиями

факт соответствует биологическим данным [10] — нейрон с наибольшим внешним воздействием становится центральным элементом в сети, таким образом, вся сеть настраивается на восприятие наиболее сильного внешнего воздействия, а остальные игнорируются. Согласно [7] возврат гистерезисной природы МРК, а значит, и возможности запоминания информации, происходит во время сна. Таким образом, моделирование нейронной сети (7) при условии  $k_i = \text{const}$ ,  $0 < k_i < 1$  может характеризовать работу мозга во сне. Согласно [13] основная задача мозга во сне — консолидация полученной информации, во время бодрствования и при достаточно сильном раздражителе запомнится именно она.

Динамику нейронной сети (7), состоящей из трех нейронов при внешних воздействиях

$$\begin{aligned} g_1 &= 1; \\ 0,4 &\leq g_2 \leq 1; \\ 0,4 &\leq g_3 \leq 1 \end{aligned} \quad (17)$$

иллюстрирует сечение Пуанкаре фазовых траекторий с плоскостью  $u_1 = 0,4$ .

Результат приведен на рис. 10 для  $t > 10$ .

Из рис. 10 видно, что  $u_2(t) \approx u_3(t)$ . Подобное поведение можно трактовать как появление в сети центрального элемента. Отметим, что подробное изучение биологических аспектов сетей с центральным элементом приведено в работе [11].

С. А. Сакулин, канд. техн. наук, доц., e-mail: sakulin@bmstu.ru,  
 А. Н. Алфимцев, канд. техн. наук, доц., e-mail: alfim@bmstu.ru  
 Московский государственный технический университет им. Н. Э. Баумана

## Формализация экспертных знаний об удобстве веб-страниц на основе агрегирования пользовательских критериев

*Интеграл Шоке по нечеткой мере, являющийся обобщением средневзвешенного оператора, за счет своей гибкости в агрегировании информации, а также из-за способности учитывать возможные неопределенности в данных получил широкую популярность в приложениях многокритериального принятия решений. Рассматриваются его применение в рамках методики формализации экспертных знаний об удобстве веб-страниц и решение задачи идентификации нечеткой меры для агрегирования пользовательских критериев. Приводятся примеры оценок для девяти основных критериев удобства веб-страниц, а также результаты оценки веб-страниц сайтов ведущих университетов мира.*

**Ключевые слова:** юзабилити, оператор агрегирования, нечеткая мера, интеграл Шоке

S. A. Sakulin, A. N. Alfimtsev

## Formalization of Expert Knowledge about the Usability of Web Pages Based on User Criteria Aggregation

*The Choquet integral with respect to fuzzy measure is a generalization of weighted arithmetic mean aggregation operator. Due to its flexibility in the aggregation of information and because of the ability to take into account the possible uncertainties in the data has become popular in the applications of multi-criteria decision making. In this paper we consider its use in the formalization of expert knowledge about the convenience of Web pages. A solution of the fuzzy measure identification problem to aggregate user-defined criteria is discussed. The evaluation examples for the nine main criteria and the evaluation of world's leading university's Web pages is considered.*

**Keywords:** usability, aggregation operator, fuzzy measure, Choquet integral

### Введение

Всего лишь пятнадцать лет назад для принятия какого-либо решения в любой сфере (выбор места проведения отпуска, покупка товара, методы лечения заболеваний и т. п.) индивидуум полагался, как правило, на личный опыт и на информацию, взятую из различных печатных источников, а также на опыт родителей, друзей, но в начале нового века основным источником такой информации де факто стал Интернет [1]. В глобальной сети можно найти как веб-страницы, содержащие экспертную информацию разной степени компетентности, так и большое число электронных версий печатных источников различного качества. В связи с этим все острее встает вопрос профессиональной оценки удобства использования и достоверности подобных источников информации в глобальной сети. Вместе с тем, сам пользователь может сформулировать критерии и способ их агрегирования для создания обобщенного критерия оценки того или иного свойства источников информации, чтобы выбрать источник на основе своих предпочтений.

Таким образом, для оценки удобства использования информационных источников представля-

ется актуальным вопрос формализации связанных с такой оценкой экспертных знаний. В настоящей статье рассмотрен вопрос формализации экспертных знаний, касающихся оценки удобства использования веб-страниц посредством агрегирования пользовательских критериев с помощью нечеткого оператора агрегирования Шоке (интеграла Шоке).

### Критерии оценки удобства веб-страниц

Пользовательские критерии оценки удобства веб-страниц можно сформулировать по-разному, в зависимости от целей проводимой оценки и от конкретного эксперта. Эти критерии условно делятся на несколько групп [2]. Из этих групп критериев, в свою очередь, в соответствии с логикой эксперта можно построить определенную иерархию. Рассмотрим подобную иерархию и каждый из критериев пользователя по отдельности, основываясь на работе [3].

Для того чтобы достигнуть простого и удобного просмотра веб-страницы, она должна удовлетворять нескольким критериям.

Во-первых, это ясность визуальной иерархии страницы для пользователя, обозначим ее  $G_1$ . Она

включает в себя следующие несколько составляющих (входных критериев):  $g_1^1$  — степень "выделенности" или "подчеркнутости" наиболее важных элементов на странице;  $g_2^1$  — степень визуальной связанности элементов, которые связаны между собой логически;  $g_3^1$  — оценка выполнения условия: "если элементы являются частями друг друга логически, то они должны быть представлены в виде вложений".

Во-вторых, степень использования на странице принятых и устоявшихся условностей, общих правил, обозначим этот критерий  $G_2$ . Критерий состоит из двух входных критериев:  $g_1^2$  — применение стандартных обозначений и условностей;  $g_2^2$  — ясное обозначение активных элементов.

В-третьих, мера разделенности веб-страницы на четкие области для того, чтобы пользователю было удобно в этих областях ориентироваться, обозначим ее  $G_3$ . Критерий включает в себя следующие составляющие:  $g_1^3$  — применение шаблона "поиск, категории, содержание";  $g_2^3$  — использование сетчатого макета.

В-четвертых, степень наличия визуального шума, обозначим ее  $G_4$ . Этот критерий включает следующие составляющие:  $g_1^4$  — мера "перегруженности" страницы элементами;  $g_2^4$  — степень наличия фонового шума.

На рис. 1 на примерах известных веб-страниц проиллюстрированы каждый из перечисленных критериев.

Рассмотрим пример качественных рассуждений эксперта относительно агрегирования приведенных критериев. Очевидно, что некоторые из этих критериев коррелированы, в частности  $G_1$  и  $G_2$ , поскольку интуитивно понятно, что если веб-страница обладает свойством ясности визуальной иерархии, то, скорее всего, разработчик позаботился о хорошем разделении ее содержимого на четкие области. Аналогично, критерии  $g_1^2$  и  $g_2^2$  коррелированы, поскольку ясно, что если веб-страница создана с применением стандартных обозначений и условностей, то, скорее всего, на ней ясно обозначены активные элементы.

В соответствии с качественными рассуждениями эксперта, если веб-страница в достаточной степени обладает свойством  $g_1^1$ , т. е. наиболее важные элементы хорошо заметны и выделены, то в таких условиях критерий  $g_3^1$  становится более важным для оценки ясности визуальной иерархии  $G_1$ , чем критерий  $g_2^1$ , поскольку при этом гиперссылки "наиболее важный элемент —

его составные части" будут более удобны для пользователя, чем составные части визуально связанные между собой, но расположенные рядом с выделенным наиболее важным элементом. Это явление носит название *предпочтительная зависимость критериев*. Известно [4], что никакой аддитивный оператор агрегирования, в том числе средневзвешенный, не позволяет формализовать подобные рассуждения эксперта. Кроме того, эксперт отмечает, что при определении степени наличия визуального шума  $G_4$  более важной является  $g_1^4$  — мера "перегруженности" страницы элементами, чем  $g_2^4$  — степень наличия фонового шума.

Рассуждения экспертов подобного вида могут быть достаточно сложными и неоднозначными, зависящими от мнения конкретного эксперта. В силу этого обстоятельства формализация экспертных знаний должна позволять экспертам понимать и корректировать рассуждения друг друга, чтобы в

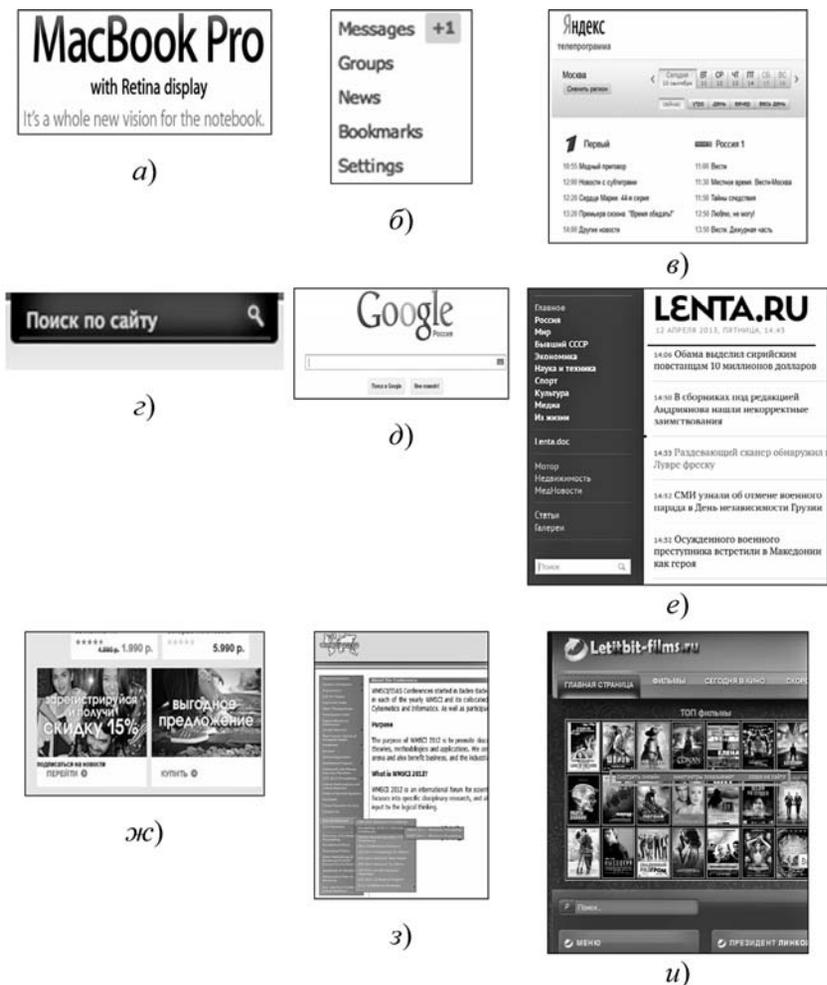


Рис. 1. Примеры достижения ясного и удобного просмотра веб-страниц:

$a$  —  $g_1^1$  — apple.com;  $b$  —  $g_2^1$  — vk.com;  $v$  —  $g_3^1$  — tv.yandex.ru;  $z$  —  $g_1^2$  — progbook.com;  $d$  —  $g_2^2$  — google.ru;  $e$  —  $g_1^3$  — lenta.ru;  $ж$  —  $g_2^3$  — adidas.ru;  $з$  —  $g_1^4$  — iiis-summer13.org;  $u$  —  $g_2^4$  — letitbit-films.ru

итоге получить согласованную формальную модель. Для этого необходимо, чтобы модель позволяла одному эксперту проследить всю цепочку рассуждений другого эксперта и скорректировать ее в соответствии со своим представлением, обосновав при этом все свои корректировки. В результате такой совместной работы экспертов можно получить согласованную модель оценки. Процесс ее создания будет итеративным, каждая итерация при этом будет представлять собой "внесение обоснованных поправок" тем или иным экспертом.

### Операторы агрегирования критериев оценки удобства веб-страниц

Оператором агрегирования часто называют обладающую некоторыми заданными свойствами функцию от  $H$  переменных (критериев), каждая из которых определена на единичном интервале. Областью значений этой функции также является единичный интервал. В соответствии с этим оператор агрегирования обозначают  $AGG: [0, 1]^H \rightarrow [0, 1]$ , где  $H$  — число критериев [5]. Таким образом, для формализации экспертных знаний, рассмотренных в предыдущем разделе, необходимо найти подходящий оператор  $AGG(g_1^1, \dots, g_3^1, g_1^2, g_2^2, g_1^3, g_2^3, g_1^4, \dots, g_3^4)$ . Результатом агрегирования с помощью этого оператора будет оценка  $\Omega$  удобства веб-страницы в интервале  $[0, 1]$ . С помощью этой оценки можно будет ранжировать страницы по их удобству. В этом смысле более важна разность в оценке для отдельно взятых страниц, чем абсолютные значения этой оценки.

Для работы с критериями на единичной шкале необходима нормализация всех критериев. Например, если оценка степени соответствия веб-страницы тому или иному критерию выставляется в соответствии с десятибалльной шкалой  $\{0, 1, \dots, 9, 10\}$ , то нормализация этой шкалы даст в результате шкалу  $\{0, 0.1, \dots, 0.9, 1\}$  и будет состоять в делении каждой возможной оценки на 10. В случае если для разных критериев оценки используются разные шкалы, все эти шкалы необходимо свести к единичному отрезку (нормализовать).

Выбор того или иного оператора обусловлен информацией об агрегировании критериев, которую можно получить от эксперта. Самым простым и наиболее распространенным на практике оператором агрегирования критериев традиционно является средневзвешенный оператор. Этот оператор, однако, не позволяет формализовать знания эксперта о зависимостях между критериями [4]. Наиболее подходящим с практической точки зрения представляется применение интеграла Шоке 2-го порядка, поскольку он позволяет формализовать рассуждения эксперта, подобные описанным в предыдущем разделе, оставаясь при этом достаточно простым. В работе [6] рассмотрены вопросы применения этого аппарата в различных практических областях. В соответствии с иерархией рассмотрен-

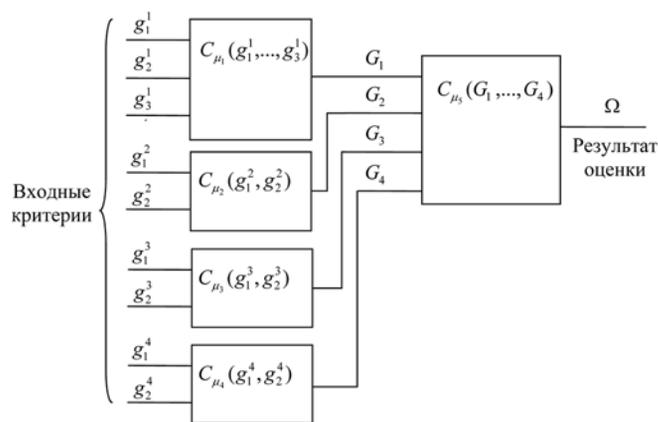


Рис. 2. Дерево из интегралов Шоке для оценки удобства веб-страницы

ных критериев можно построить дерево из интегралов Шоке 2-го порядка, подобное дереву, описанному в работах [7, 8]. Соответствующая структура представляет собой "вложенные" друг в друга интегралы Шоке (рис. 2).

На рис. 2 показаны все перечисленные в предыдущем разделе входные критерии, а также составные критерии, которые сами являются результатами агрегирования входных критериев. Интегралы Шоке  $C_{\mu_1}(g_1^1, \dots, g_3^1)$ ,  $C_{\mu_2}(g_1^2, g_2^2)$ ,  $C_{\mu_3}(g_1^3, g_2^3)$ ,  $C_{\mu_4}(g_1^4, g_2^4)$  служат для получения значений составных критериев  $G_1, \dots, G_4$ . Интеграл  $C_{\mu_5}(G_1, \dots, G_4)$  служит для агрегирования всех составных критериев. Результатом оценки является значение этого интеграла, полученное на основе значений входных критериев для конкретной веб-страницы. Выразим результат оценки  $\Omega$  через соответствующие интегралы Шоке:

$$\Omega = C_{\mu_5}(C_{\mu_1}(g_1^1, \dots, g_3^1), C_{\mu_2}(g_1^2, g_2^2), C_{\mu_3}(g_1^3, g_2^3), C_{\mu_4}(g_1^4, g_2^4)). \quad (1)$$

### Пример формализации экспертных знаний об удобстве веб-страниц

Сегодня все большей популярностью пользуются университетские сайты, к которым обращаются несколько целевых аудиторий: абитуриенты для получения информации о поступлении, студенты для ориентации в университетских событиях и преподаватели, которые ищут справочную информацию. Например, по статистике информационного центра МГТУ им. Н. Э. Баумана к сайту [bmstu.ru](http://bmstu.ru) обращаются каждый день до 15 тыс. пользователей. Соответственно любое удобство или неудобство, вызванное у пользователя при взаимодействии с сайтом, приводит к переводу этих эмоций на оценку университета в целом для тысяч людей ежедневно.

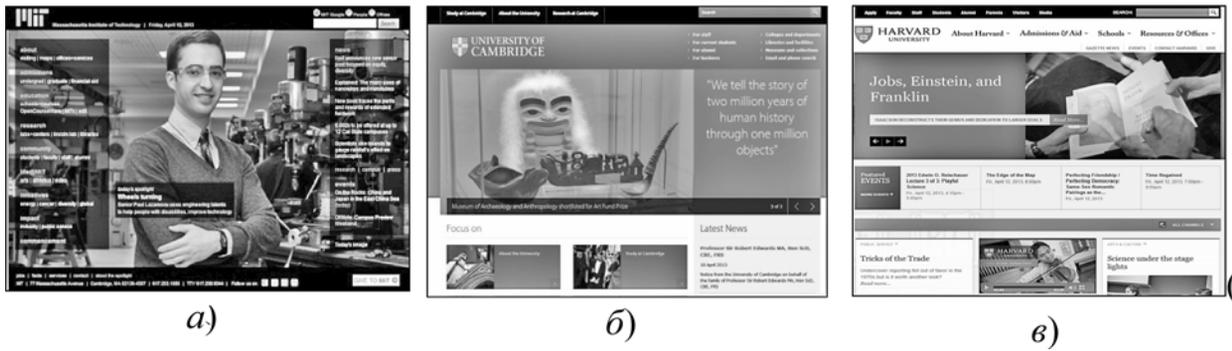


Рис. 3. Начальные веб-страницы сайтов ведущих университетов мира: а — МТИ — web.mit.edu; б — КУ — cam.ac.uk; в — ГУ — harvard.edu

На порядок это значение увеличивается для сайтов ведущих университетов мира. Поэтому рассмотрим по представленным критериям удобство начальных веб-страниц трех ведущих университетов (по рейтингу QS): Массачусетский технологический институт (МТИ) — web.mit.edu, Кембриджский университет (КУ) — cam.ac.uk и Гарвардский университет (ГУ) — harvard.edu (рис. 3), а также проведем формализацию оценки удобства этих сайтов.

Для формирования статистически репрезентативной выборки по алгоритму стратификации были выбраны оценки веб-страниц данных сайтов по десятибалльной шкале 45 студентов, обучающихся на старших курсах кафедры ИУЗ МГТУ им. Н. Э. Баумана. В соответствии с этим алгоритмом выборка оценок состояла из оценок, составленных студентами, которые были отобраны следующим образом. Из учебных групп студентов был составлен общий список. После проверки на ошибки и отсутствие повторяемости был выбран каждый третий элемент списка. Затем каждый выбранный студент провел оценку, сформировав тем самым выборку результирующих оценок. Результаты оценки, которые проводились по основным правилам тестирования юзабилити [2], представлены в табл. 1.

Несмотря на лидерство среди университетов первый же критерий  $g_1^1$  для МТИ оказался оценен очень низко, а для КУ и ГУ степень "выделенности"

наиболее важных элементов на странице оказалась вполне приличной. Степень визуальной связанности элементов  $g_2^1$  была оценена вполне удовлетворительно для всех сайтов. Вложенность элементов  $g_3^1$  не является определяющим критерием для данного типа сайтов, поэтому ее значения средние. Применением стандартных обозначений и условностей  $g_1^2$  отличились также все три университета, конечно, с учетом тех специальностей, на которые они ориентированы. Несмотря на визуальную схожесть используемых приемов построения веб-страниц критерий ясного обозначения активных элементов  $g_2^2$  был оценен почти максимально у ГУ и КУ, немного меньше у МТИ.

Все сайты использовали шаблон пользовательского интерфейса: "поиск, категории, содержание", но именно у МТИ, критерий  $g_1^3$  имел относительно малое значение, возможно, из-за неудачного цветового решения, хотя напрямую критерий "цвет" здесь не рассматривался. Можно отметить, что в целом критерий разделенности веб-страницы на четкие области и, в частности  $g_2^3$ , имел максимальные значения у сайта КУ. Наиболее перегруженной страницей  $g_1^4$  являлась веб-страница МТИ. Кроме того, не удовлетворила экспертов степень наличия фонового шума  $g_2^4$  на сайте технологического института.

Перед экспертами также ставился вопрос общей оценки удобства каждой из веб-страниц. Для экспертов было очевидным, что сайты ГУ и КУ более удобны, чем сайт МТИ, что можно формализовать в виде следующих отношений предпочтения:

$$g_2 \succ_{\mu_5} g_1; \quad (2)$$

$$g_3 \succ_{\mu_5} g_1. \quad (3)$$

Здесь  $g_1, \dots, g_3$  — реализации входных критериев для сайтов МТИ, КУ и ГУ соответственно. Индекс  $\mu_5$  означает, что отношения предпочтения рассматриваются в рамках итоговой оценки, которая формируется интегралом  $G_{\mu_5}(G_1, \dots, G_4)$ . Другие предпочтения не так очевидны для экспертов. С учетом описанных выше качественных рассуждений одного из экспертов относительно влияния на результат

Таблица 1

Значения критериев для рассматриваемых сайтов

Входной критерий	МТИ ( $g_1$ )	КУ ( $g_2$ )	ГУ ( $g_3$ )
$g_1^1$	0,3	0,9	0,9
$g_2^1$	0,7	0,7	0,5
$g_3^1$	0,4	0,5	0,7
$g_1^2$	0,6	0,6	0,6
$g_2^2$	0,6	0,9	0,9
$g_1^3$	0,4	0,8	0,9
$g_2^3$	0,4	0,9	0,8
$g_1^4$	0,2	0,8	0,8
$g_2^4$	0,3	0,8	0,9

критериев  $g_1^1, \dots, g_3^1$ , эксперты пришли к общему выводу о том, что оценка  $\Omega$  для сайта ГУ должна быть выше оценки для сайта КУ, что можно выразить отношением предпочтения:

$$g_3 \succ_{\mu_5} g_2. \quad (4)$$

Положительная корреляция критериев  $G_1$  и  $G_2$ , а также положительная корреляция критериев  $g_1^2$  и  $g_2^2$  формализуется с помощью знака соответствующих индексов взаимодействия критериев:

$$I_{\mu_5}(1,2) < 0; \quad (5)$$

$$I_{\mu_2}(1,2) < 0. \quad (6)$$

Здесь индексы  $\mu_5$  и  $\mu_2$  означают, что индексы взаимодействия критериев относятся к интегралам Шоке по нечеткой мере  $\mu_5$  и  $\mu_2$  соответственно. Приведенные выше рассуждения эксперта о том, что при определении степени наличия визуального шума  $G_4$  более важной является  $g_1^4$  — мера "перегруженности" страницы элементами, чем  $g_2^4$  — степень наличия фонового шума, формализуются с помощью соответствующих индексов Шепли:

$$\Phi_{\mu_4}(1) > \Phi_{\mu_4}(2). \quad (7)$$

Здесь индекс  $\mu_4$  означает, что индекс Шепли относится к интегралу Шоке по нечеткой мере  $\mu_4$ . Кроме того, при оценке ясности визуальной иерархии эксперт указал на предпочтительную зависимость критериев  $g_1^1, \dots, g_3^1$ . Это явление формализуется следующим отношением предпочтения:

$$g_3 \succ_{\mu_1} g_2. \quad (8)$$

Индекс  $\mu_1$  здесь показывает, что отношение предпочтения рассматривается в рамках локальной оценки критерия  $G_1$ , формируемой интегралом  $C_{\mu_1}(g_1^1, \dots, g_3^1)$ .

### Идентификация нечетких мер

Для того чтобы получить результат в соответствии с формулой (1), необходимо идентифицировать нечеткие меры  $\mu_1, \dots, \mu_5$  на основе экспертных предпочтений, выраженных ограничениями (2)—(8). Для идентификации был выбран метод максимизации энтропии нечеткой меры при заданных ограничениях. Этот метод основан на принципе, предложенном Джейнсом. Принцип Джейнса [8] заключается в том, что если мы обладаем какой-то частью знаний о поведении случайной величины, то к той части знаний, которая нам недоступна, следует относиться наиболее непредвзято, максимизируя энтропию этой величины с учетом имеющихся знаний. В отношении идентификации нечетких мер этот принцип предложил применять Кожудинович [9]. В результате мы получим наименее специфичную нечеткую меру из всех возможных при ограничениях, обусловленных экспертными

предпочтениями. В соответствии с методом максимизации энтропии [9, 10] отношения предпочтения (2)—(4) переводятся в неравенства:

$$\begin{aligned} & C_{\mu_5}(C_{\mu_1}(0,9, 0,7, 0,5), C_{\mu_2}(0,6, 0,9), \\ & C_{\mu_3}(0,8, 0,9), C_{\mu_4}(0,8, 0,8)) - \\ & - C_{\mu_5}(C_{\mu_1}(0,3, 0,7, 0,4), C_{\mu_2}(0,6, 0,6), \\ & C_{\mu_3}(0,4, 0,4), C_{\mu_4}(0,2, 0,3)) \geq \delta_C^{\mu_5}; \quad (9) \end{aligned}$$

$$\begin{aligned} & C_{\mu_5}(C_{\mu_1}(0,9, 0,5, 0,7), C_{\mu_2}(0,6, 0,9), \\ & C_{\mu_3}(0,9, 0,8), C_{\mu_4}(0,8, 0,8)) - \\ & - C_{\mu_5}(C_{\mu_1}(0,3, 0,7, 0,4), C_{\mu_2}(0,6, 0,6), \\ & C_{\mu_3}(0,4, 0,4), C_{\mu_4}(0,2, 0,3)) \geq \delta_C^{\mu_5}; \quad (10) \end{aligned}$$

$$\begin{aligned} & C_{\mu_5}(C_{\mu_1}(0,9, 0,5, 0,7), C_{\mu_2}(0,6, 0,9), \\ & C_{\mu_3}(0,9, 0,8), C_{\mu_4}(0,8, 0,8)) - \\ & - C_{\mu_5}(C_{\mu_1}(0,9, 0,7, 0,5), C_{\mu_2}(0,6, 0,9), \\ & C_{\mu_3}(0,8, 0,9), C_{\mu_4}(0,8, 0,8)) \geq \delta_C^{\mu_5}. \quad (11) \end{aligned}$$

Неравенства (5)—(7) преобразуются, соответственно, в неравенства:

$$-1 \leq I_{\mu_5}(1,2) \leq -\delta_I^{\mu_5}; \quad (12)$$

$$-1 \leq I_{\mu_2}(1,2) \leq -\delta_I^{\mu_2}; \quad (13)$$

$$\Phi_{\mu_4}(1) - \Phi_{\mu_4}(2) \geq \delta_{\Phi}^{\mu_4}. \quad (14)$$

Отношение предпочтения (8) переводится в неравенство:

$$C_{\mu_1}(0,9, 0,5, 0,7) - C_{\mu_1}(0,9, 0,7, 0,5) \geq \delta_C^{\mu_1}. \quad (15)$$

Здесь  $\delta_C^{\mu_1}, \delta_C^{\mu_5}, \delta_I^{\mu_2}, \delta_I^{\mu_5}, \delta_{\Phi}^{\mu_4}$  — задаваемые экспертом пороги безразличия для идентификации соответствующих нечетких мер. Исходя из того, что все входные критерии определены на шкалах вида  $\{0,0,1, \dots, 0,9,1\}$ , а также принимая во внимание ограничения на значения порогов безразличия, применяемые для недопущения выбора значений порогов, при которых заведомо не имеет решения задача идентификации нечеткой меры, для мер  $\mu_1, \dots, \mu_4$  были выбраны следующие значения этих порогов:  $\delta_C^{\mu_1} = 0,055, \delta_I^{\mu_2} = 0,005, \delta_{\Phi}^{\mu_4} = 0,002$ . Для меры  $\mu_5$  с учетом ограничений, рассмотренных в работе [10], и неравенств (9)—(12) были выбраны значения порогов  $\delta_C^{\mu_5} = 0,03, \delta_I^{\mu_5} = 0,0027$ .

Параметры и значения интегралов Шоке

Интегралы Шоке	Параметры									
	$\Phi(1)$	$\Phi(2)$	$\Phi(3)$	$\Phi(4)$	$I(1,2)$	$I(2,3)$	$I(1,3)$	$g_1$	$g_2$	$g_3$
$C_{\mu_1}$	0,3333	0,2474	0,4193	—	-0,10313	0	0,10313	0,4564	0,6725	0,7275
$C_{\mu_2}$	0,5	0,5	—	—	-0,005	—	—	0,6	0,7507	0,7507
$C_{\mu_3}$	0,5	0,5	—	—	0	—	—	0,4	0,85	0,85
$C_{\mu_4}$	0,51	0,49	—	—	0	—	—	0,249	0,8	0,849
$C_{\mu_5}$	0,2677	0,2341	0,2341	0,2640	-0,0027	-0,004	0,0295	0,4254	0,7638	0,7939

Для идентификации нечетких мер  $\mu_1, \dots, \mu_5$  был использован специализированный пакет Kappalab [12]. На первом этапе были идентифицированы меры  $\mu_1, \dots, \mu_4$  на основе ограничений (13)—(15). Результаты этой идентификации в виде параметров соответствующих интегралов (индексов взаимодействия и индексов Шепли), а также значения этих интегралов на реализациях  $g_1, g_2, g_3$  приведены в табл. 2 в соответствующих столбцах. На втором этапе также с помощью метода максимизации энтропии была идентифицирована нечеткая мера  $\mu_5$  на основе накладываемых на нее ограничений (9)—(12) и полученных на первом этапе значений интегралов  $C_{\mu_1}, \dots, C_{\mu_4}$ . В табл. 2 в нижней строке приведены параметры и значения интеграла  $C_{\mu_5}$  на реализациях  $g_1, g_2, g_3$ .

Как и ожидалось, эти значения (результаты агрегирования) отражают экспертные оценки удобства соответствующих веб-страниц, сформулированные в виде отношений предпочтения (2)—(4) и переведенные затем в неравенства (9)—(11). Отмеченное экспертом отрицательное взаимодействие (корреляция) критериев  $G_1$  и  $G_2$  (5) выражено отрицательным индексом взаимодействия  $I_{\mu_5}(1,2)$ . Корреляция критериев  $g_1^2$  и  $g_2^2$  (6) выражена отрицательным индексом взаимодействия  $I_{\mu_2}(1,2)$ . Различие в субъективном "весе" критериев  $g_1^4$  и  $g_2^4$  (7) отражено соответствующими значениями индексов Шепли  $\Phi_{\mu_4}(1)$  и  $\Phi_{\mu_4}(2)$ . Предпочтительная зависимость критериев  $g_1^1, \dots, g_3^1$  отражена в результатах идентификации значениями интеграла  $C_{\mu_1}$  на реализациях  $g_1, g_2, g_3$ . Таким образом, при идентификации нечетких мер  $\mu_1, \dots, \mu_5$  были отражены все экспертные предпочтения. Коэффициенты этих мер служат параметрами для соответствующих интегралов, которые, в свою очередь, образуют дерево оценки.

## Заключение

В человеко-машинном взаимодействии можно выделить две составляющие: человеческую и машинную. И несмотря на значительное число философских, культурологических и социально-психологических теорий именно человеческая составляющая остается наиболее сложной для изучения. В предположении о том, что методы технического университета, заключающиеся в формальном описании процессов, в частности, человеко-машинного взаимодействия с веб-страницами, будут действенным средством в решении этой задачи, в статье рассмотрена методика формализации экспертных знаний об удобстве веб-страниц. В продолжение данной работы планируется расширить статистику экспериментального исследования за счет автоматизации указанной методики в качестве приложения университетского сайта.

*Работа выполнена при поддержке Стипендии Президента РФ СП-687.2013.5.*

## Список литературы

1. Barber N. A., Paul P. T. Investigating the potential influence of the Internet as a new socialization agent in context with other traditional socialization agents // Journal of marketing theory and practice. 2013. N 2. P. 179—194.
2. Скотт Б., Нейл Т. Проектирование веб-интерфейсов. СПб.: Символ-Плюс, 2010. 352 с.
3. Круг С. Как сделать сайт удобным. Юзабилити по методу Стива Круга. СПб: Питер, 2010. 208 с.
4. Grabisch M., Orlovski S., Yager R. Fuzzy aggregation of numerical preferences // Handbook of Fuzzy Sets Series / R. Slowinski (ed.), Dordrecht: Kluwer Academic, 1998. Vol. 4. P. 31—68.
5. Detyniecki M. Mathematical Aggregation Operators and their Application to Video Querying: Thesis for the degree Docteur de l'Université. Paris, 2000. 185 p.
6. Сакулин С. А., Алфимцев А. Н. К вопросу о практическом применении нечетких мер и интеграла Шоке // Инженерный журнал: наука и инновации. 2012. Вып. 1. URL: <http://engjournal.ru/catalog/it/hidden/71.html>
7. Narukawa Y., Torra V. Twofold integral and Multi-step Choquet integral // Kybernetika-Praha. 2004. N 40 (1). P. 39—50.
8. Тимонин М. В. Проблема оптимизации распределения ресурсов при планировании стратегии информационной безопасности // Безопасность информационных технологий. 2011. N 2. С. 90—96.
9. Jaynes E. T. Information theory and statistical mechanics // Phys. Rev. 1957. N 106. P. 620—630.
10. Kojadinovic I. Minimum variance capacity identification // European Journal of Operational Research. 2007. N 177 (1). P. 498—514.
11. Grabisch M., Kojadinovic I., Meyer P. A review of methods for capacity identification in Choquet integral based multi-attribute utility theory // Applications of the Kappalab R package. 2008. N 2. P. 766—785.

А. А. Кузьмин, студент, А. А. Адуенко, студент, Московский физико-технический институт,  
В. В. Стрижов, канд. физ.-мат. наук, науч. сотр., Вычислительный Центр РАН, e-mail: strijov@ccas.ru

## Тематическая классификация тезисов крупной конференции с использованием экспертной модели

*Работа посвящена определению тем, научных направлений и сессий тезисов крупной научной конференции. Рассматривается коллекция тезисов конференции с экспертной тематической моделью. Строится терминологический словарь конференции. Предлагается функция сходства двух тезисов. Методом неметрической иерархической кластеризации строится алгоритмическая модель конференции, с заданным весом учитывающая существующую экспертную модель. Выявляются несоответствия между экспертной моделью и предлагаемой. Алгоритм построения тематической модели проиллюстрирован кластеризацией коллекции тезисов конференции EURO 2013.*

**Ключевые слова:** коллекция документов, тематические модели, иерархические модели, кластеризация

A. A. Kuzmin, A. A. Aduenko, V. V. Strijov

## Thematic Classification Using Expert Model for Major Conference Abstracts

*This paper is devoted to the thematic verification of the areas, streams and sessions of a major conference. An abstract collection and its expert thematic model are considered. The terminological dictionary of the conference is constructed. A similarity function between two abstracts is proposed. A non-metric clustering algorithm is used to construct the hierarchical thematic model of the conference. The expert model is considered in this algorithm and its parameters are optimized, to make the algorithmic model similar to the expert model. The expert model is compared with algorithmic model. The algorithm is illustrated by clustering of the EURO 2013 abstracts.*

**Keywords:** document collection, thematic model, hierarchical model, clustering

### Введение

При организации крупной конференции возникает задача построения ее тематической модели. Рассмотрим процесс построения тематической модели тезисов (далее документов) на примере конференции EURO 2013. Документом является аннотация к докладу участника конференции, состоящая не более чем из 600 символов. Конференция содержит в себе 24 главные темы (далее область), определяемые председателем программного комитета. Каждая главная тема содержит в себе примерно 10 больших подтем (далее направление), каждая из которых делится на сессии, содержащие в себе четыре документа. Эксперты, исходя из своей стратегии организации докладов конференции и основываясь на содержании документа, относят его к одной из областей, затем к одному из направлений данной области. После этого все документы, попавшие в одно направление, разбиваются на сессии.

В силу большого числа экспертов и отсутствия эталонной модели оценить качество экспертной тематической модели сложно. Поэтому предлагается построить алгоритмическую тематическую модель

коллекции документов, основанную на их терминологическом сходстве, и сравнить ее с экспертной.

Для этого сначала составляют экспертным образом терминологический словарь конференции и из всех документов удаляют слова, не являющиеся терминами. После отсева неинформативных слов документы представляют в виде "мешков слов" [1] и каждому документу ставится в соответствие целочисленный вектор. В работе [2] сравниваются способы построения вектора — описания документа.

Кластеризация текстов выполняется с помощью метрических алгоритмов кластеризации, например, K-means [3], FOREL [4], C-means [5], STOLP [6], FRiS-STOLP [7], BoostML, DANN [8] и других [9, 10], и с помощью вероятностных методов [11], например, вероятностного латентного семантического анализа [1] или латентного размещения Дирихле [12].

Для построения иерархической кластерной модели существуют два основных типа алгоритмов — дивизимные и агломеративные [13]. В данной работе для построения тематической модели предлагается неметрический алгоритм кластеризации (метрический алгоритм для аналогичной задачи был предложен в работе [13]), перераспределяющий объекты между кластерами так, чтобы улучшить

среднее сходство между объектами из одного кластера и увеличить среднее различие между объектами из разных кластеров. Требуется построить иерархическую модель, сохранив ее схожесть с экспертной, поэтому при кластеризации учитывается модель, предложенная экспертами, с определенным весом. При построении иерархической модели проводится такая кластеризация, согласно которой объект принадлежит только к одному из кластеров, так как каждый документ может принадлежать только одной теме, что соответствует правилам проведения конференции.

## 1. Математическое представление конференции

Пусть  $W = \{w_1, \dots, w_n\}$  — заданное множество слов (словарь), где  $n$  — число слов в словаре. Документом  $d$  из коллекции  $D$  назовем неупорядоченное множество слов из  $W$ ,  $d = \{w_j, j \in \{1, \dots, n\}\}$ .

Поставим в соответствие каждому документу  $d$  его описание — вектор  $\mathbf{x}$  размерности  $n$  следующим образом: если слово  $w_j$  из словаря  $W$  встретилось в документе  $d_s$   $k$  раз, то  $x_{s,j} = k$ ,  $k \geq 0$ . Получим матрицу  $\mathbf{X}$  — "объект-признак", где каждая строка  $\mathbf{x}_s = [x_{s,1}, \dots, x_{s,n}]$  — признаковое описание документа  $d_s$ :

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & \dots & x_{1,n} \\ \dots & \dots & \dots \\ x_{|D|,1} & \dots & x_{|D|,n} \end{pmatrix}. \quad (1)$$

Для удобства дальнейшего изложения нормируем все строки матрицы  $\mathbf{X}$  следующим образом:

$$\mathbf{x}_s \mapsto \frac{\mathbf{x}_s}{\sqrt{\mathbf{x}_s^T \mathbf{x}_s}}. \quad (2)$$

Представим иерархическую тематическую модель в виде дерева (рис. 1). Глубину дерева обозначим  $h$ , на рис. 1 глубина  $h = 5$ . Уровнем  $l$  иерархии назовем множество всех узлов дерева, находящихся на глубине  $l$ . Документы  $d_s \in D$  являются листьями этого дерева и имеют уровень  $h$ . Кластером  $c$  будем называть подмножество коллекции доку-

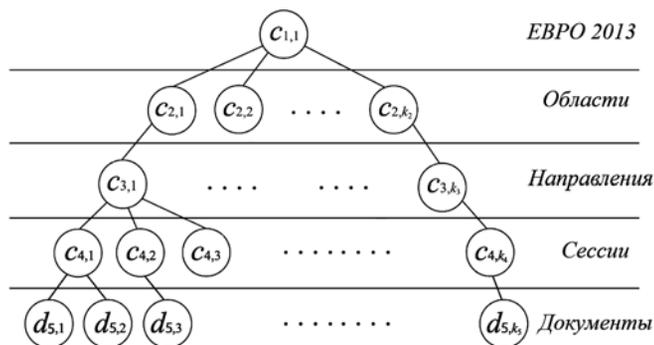


Рис. 1. Иерархическое представление тематической модели

ментов  $D$ . Поставим в соответствие каждому узлу  $i$  уровня  $l$  дерева кластер  $c_{l,i}$ , состоящий из документов  $d_s$ , путь до которых от вершины  $c_{1,1}$  проходит через узел  $(l, i)$ .

## 2. Функция сходства документов

Предполагается, что каждый документ в коллекции может быть описан небольшим набором признаков — ключевых слов. В рассматриваемой в данной работе коллекции [14] каждый документ описывается 10—15 признаками. При этом словарь состоит из более 1000 слов.

Предлагается ввести функцию сходства  $s(\mathbf{x}_i, \mathbf{x}_j)$  двух документов:

$$s(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathbf{x}_i^T \mathbf{x}_j}{\sqrt{\mathbf{x}_i^T \mathbf{x}_i} \sqrt{\mathbf{x}_j^T \mathbf{x}_j}} = \mathbf{x}_i^T \mathbf{x}_j \quad (3)$$

В функции (3) учтена нормировка (2), позволяющая документам  $\mathbf{x}_i$  и  $\mathbf{x}_j$  иметь разную длину в словах при сравнении. Так как все компоненты векторов  $\mathbf{x}_i, \mathbf{x}_j$  неотрицательны, то  $s(\mathbf{x}_i, \mathbf{x}_j) \in [0, 1]$ , причем  $s = 1$  достигается для документов, словарный состав которых одинаков.

Под сходством  $S(c_{l,i}, c_{l,j})$  двух кластеров  $c_{l,i}$  и  $c_{l,j}$  уровня  $l$  будем понимать среднее сходство  $s(\mathbf{x}, \mathbf{y})$  между документами  $\mathbf{x} \in c_{l,i}, \mathbf{y} \in c_{l,j}$ , содержащимися в них. Среднее сходство  $S(\cdot, \cdot)$  внутри одного кластера для каждого документа  $d_s$  определяется как среднее сходство  $s(\cdot, \cdot)$  с остальными документами данного кластера:

$$S(c_{l,i}, c_{l,j}) = \frac{1}{|A|} \sum_{(\mathbf{x}, \mathbf{y}) \in A} s(\mathbf{x}, \mathbf{y}), \quad (4)$$

где  $A$  — множество всех пар документов из кластеров  $c_{l,i}$  и  $c_{l,j}$  таких, что  $\mathbf{x} \in c_{l,i}, \mathbf{y} \in c_{l,j}$  и  $\mathbf{x} \neq \mathbf{y}$ .

Сравним введенную функцию сходства (3) с Евклидовым расстоянием

$$\rho(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}, \quad (5)$$

с расстоянием Хеллингера

$$H(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{2}} \|\sqrt{\mathbf{x}} - \sqrt{\mathbf{y}}\|^2 \quad (6)$$

и расстоянием Дженсона—Шеннона:

$$JSD(\mathbf{x}||\mathbf{y}) = \frac{1}{2} D(\mathbf{x}||M) + \frac{1}{2} D(\mathbf{y}||M), \quad M = \frac{1}{2} (\mathbf{x} + \mathbf{y}), \quad (7)$$

где  $D(\mathbf{x}||\mathbf{y}) = \sum_i \ln \left( \frac{x_i}{y_i} \right) x_i$  — расстояние Кульбака—Лейблера.

На рис. 2,  $a-v$  (см. третью сторону обложки) приведены значения средних расстояний между

документами разных областей для указанных трех типов расстояния для экспертной кластеризации. Рис. 2, з (см. третью сторону обложки) соответствует введенной функции сходства. По осям отложены номера областей, цвет элемента  $(x, y)$  соответствует среднему расстоянию между документами области с номером  $x$  и области с номером  $y$ . Элементы диагонали  $(x, x)$  соответствуют внутрикластерному расстоянию, а элементы  $(x, y)$ ,  $x \neq y$  — межкластерному. В случае, представленном на рис. 2, з, средним расстоянием является сходство двух кластеров. Показателем качества кластеризации являются большие межкластерные расстояния по сравнению с внутрикластерными, что соответствует пику на диагонали. Отличия в среднем внутрикластерном расстоянии и среднем межкластерном расстоянии для а)–в) почти отсутствуют. Поэтому применение метрических алгоритмов кластеризации в данной задаче необоснованно. Значимые отличия между сходством или расстоянием внутри одной области и между разными областями наблюдаются только на рис. 2, з, что подтверждает целесообразность использования введенной функции сходства (4).

Определим  $\bar{x}_i$  как средний вектор в кластере  $c_{l,i}$ :

$$\bar{x}_i = \frac{1}{|c_{l,i}|} \sum_{\mathbf{x} \in c_{l,i}} \mathbf{x} \quad (8)$$

для разных областей в экспертной модели. В соответствии с (4) и (3) при  $i \neq j$

$$\begin{aligned} S(c_{l,i}, c_{l,j}) &= \frac{1}{|c_{l,i}||c_{l,j}|} \sum_{\mathbf{x} \in c_{l,i}} \sum_{\mathbf{y} \in c_{l,j}} \mathbf{x}^T \mathbf{y} = \\ &= \left( \frac{1}{|c_{l,i}|} \sum_{\mathbf{x} \in c_{l,i}} \mathbf{x} \right)^T \left( \frac{1}{|c_{l,j}|} \sum_{\mathbf{y} \in c_{l,j}} \mathbf{y} \right) = \bar{x}_i^T \bar{x}_j. \end{aligned}$$

Аналогично,

$$\begin{aligned} S(c_{l,i}, c_{l,i}) &= \frac{1}{|c_{l,i}|} \sum_{\mathbf{x} \in c_{l,i}} \frac{1}{|c_{l,i}|-1} \sum_{\mathbf{y} \neq \mathbf{x}, \mathbf{y} \in c_{l,i}} \mathbf{x}^T \mathbf{y} = \\ &= \frac{1}{|c_{l,i}|} \sum_{\mathbf{x} \in c_{l,i}} \frac{1}{|c_{l,i}|-1} \mathbf{x}^T (|c_{l,i}|\bar{x}_i - \mathbf{x}) = \\ &= \frac{1}{|c_{l,i}|} \sum_{\mathbf{x} \in c_{l,i}} \frac{|c_{l,i}|}{|c_{l,i}|-1} \mathbf{x}^T \bar{x}_i - \frac{1}{|c_{l,i}|-1} = \\ &= \frac{|c_{l,i}|}{|c_{l,i}|-1} \bar{x}_i^T \bar{x}_i - \frac{1}{|c_{l,i}|-1}. \end{aligned}$$

В последнем выражении учтена нормировка  $\mathbf{x}^T \mathbf{x} = 1$ . Таким образом, сходство и между кластерами, и внутри кластеров определяется только средними векторами кластеров, что позволяет их эффективно считать и пересчитывать при изменении состава кластеров. Введем далее функционал качества кластеризации и опишем алгоритм.

### 3. Функционал качества и алгоритм кластеризации документов

В качестве функционала качества кластеризации будем использовать комбинацию внутри- и межкластерных сходств следующего вида:

$$\begin{aligned} Q(\bar{x}_1, \dots, \bar{x}_k) &= \sum_{l=2}^{h-1} \left[ \frac{1-\alpha}{k_l} \sum_{i=1}^{k_l} |c_{l,i}| S(c_{l,i}, c_{l,i}) - \right. \\ &\quad \left. - \alpha \frac{2}{k_l(k_l-1)} \sum_{i < j} S(c_{l,i}, c_{l,j}) \right] \rightarrow \max, \quad (9) \end{aligned}$$

где  $\alpha \in [0, 1]$  — весовой коэффициент, отвечающий за приоритет при максимизации;  $k$  — общее число кластеров уровня  $l$ . При  $\alpha \rightarrow 0$  алгоритм будет максимизировать внутрикластерное сходство, вне зависимости от межкластерного, и наоборот, при  $\alpha \rightarrow 1$ . Весовой множитель  $|c_{l,i}|$  позволяет считать среднее внутрикластерное сходство не по кластерам, а по документам. Если считать среднее внутрикластерное сходство по кластерам, то возникает один кластер, собирающий множество документов, мало сходных друг с другом. Остальные кластеры мало численны и имеют высокое внутрикластерное сходство. При усреднении по документам эта особенность исчезает.

В качестве начального приближения кластеризации распределим документы  $d_s$  по кластерам  $c_{l,i}$  согласно экспертной модели. Затем на каждом шаге алгоритма по очереди выбираем по одному документу  $\mathbf{x} \in c_{l,i}$  из коллекции  $D$  и переносим его в другой кластер так, чтобы значение функционала качества  $Q$  определенного как (9) возросло. Пусть для этого его требуется перенести в кластер  $c_{l,i'}$  и эта операция дает максимальное увеличение функционала качества. Заметим, что из всех членов в сумме для  $Q$  изменяются только  $S(c_{l,i}, c_{l,i})$ ,  $S(c_{l,i}, c_{l,j})$ ,  $S(c_{l,j}, c_{l,j})$ . Новые средние векторы кластеров  $c_{l,i}$  и  $c_{l,j}$  определяются по  $\mathbf{x}$  и старым средним векторам как

$$\begin{aligned} \bar{x}_i &\rightarrow \frac{|c_{l,i}|}{|c_{l,i}|-1} \bar{x}_i - \frac{1}{|c_{l,i}|-1} \mathbf{x}, \\ \bar{x}_j &\rightarrow \frac{|c_{l,j}|}{|c_{l,j}|-1} \bar{x}_j + \frac{1}{|c_{l,j}|-1} \mathbf{x}. \end{aligned}$$

Далее вычисляем изменение функционала качества  $Q$ . Находя тот кластер  $c_{l,i'}$  в который перенесение  $\mathbf{x}$  дает наибольший эффект, переносим  $\mathbf{x}$  в  $c_{l,i'}$ , если есть улучшение. Повторяем эти шаги пока кластеризация не стабилизируется в терминах  $Q$  (9).

### 4. Построение словаря

Рассмотрим подробнее процесс составления терминологического словаря конференции  $W$ . Сначала проводится предобработка данных. После приведения слов в документах к начальной форме полу-

чается полный словарь конференции. На следующем шаге из словаря требуется исключить все слова, не являющиеся терминами. Чтобы оценить качество терминологического словаря в данной работе использовали следующий функционал:

$$D = \frac{\sum_{\mathbf{x}^i, \mathbf{x}^j \in c_{2,k}, i < j} \sum_{t=1}^n |x_t^i - x_t^j|}{\sum_{\mathbf{x}^i, \mathbf{x}^j \in c_{2,k}, i < j} \sum_{t=1}^n (x_t^i - x_t^j)}, \quad (10)$$

где  $k$  — номер произвольного кластера уровня иерархии 2. Если  $D \approx 1$ , то пересечений в терминах документов, принадлежащих одному кластеру, почти нет, это противоречит нашему предположению, что схожие документы имеют схожий терминологический состав.

При использовании критерия  $tf \cdot idf$  для отсева неинформативных слов  $D = 0,99$ . Это вызвано наличием значительного числа шумовых слов в словаре и существенным разбросом их встречаемости, поэтому отбор терминов проводился экспертно. При этом не только отбрасывались неинформативные слова, но и схожие с экспертной точки зрения термины объединялись в один. Это позволило уменьшить значение  $D$  до 0,96 и сократить словарь до 1063 терминов.

## 5. Верификация тематической модели

Для верификации тематической модели построим кластерную структуру, схожую с экспертной. Для этого модифицируем алгоритм кластеризации, описанный выше, для учета экспертной модели.

Рассмотрим операцию перенесения объекта  $x$  из одного направления в другое. Поставим в соответствие документу  $x$  пару — (область, направление), в которой на месте слова "область" стоит знак +, если у документа  $x$  область совпадает с экспертной, и знак — иначе. Аналогично, вместо слова "направление" стоит знак +, если для этого документа направление совпадает с экспертным, и знак — иначе. Например, выражение (+, —) означает, что документ находится в экспертной области, но не в экспертном направлении.

Обозначение операции переноса  $(+, +) \mapsto (+, -)$  объекта  $x$  означает, что объект  $x$  из экспертной области и направления переносится в экспертную область (остается в той же области), но в направление, не совпадающее с экспертным. Переносы вида  $(-, +) \mapsto (+, +)$  не рассматриваются, так как имеет место свойство вложенности: экспертное направление не может находиться вне экспертной области. Каждому варианту переноса будет соответствовать штраф  $\delta$  за осуществление такого переноса.

Пусть  $Q_1$  — значение оптимизируемой функции  $Q$  (9) до переноса документа  $x$ , а  $Q_2$  — ее значение

после переноса. Перенос объекта  $x$  будем теперь осуществлять только при выполнении условия:

$$Q_2 - Q_1 \geq \delta.$$

Задавая различные штрафы, мы с разным весом учитываем существующую экспертную модель. Если требуется выявить небольшое число наиболее сильных тематических противоречий, то штрафы на перемещение документа из его экспертного кластера следует задавать большие. Если же требуется построить алгоритмическую модель без учета экспертной модели, то штрафы следует устремить к нулю. В таблице приведена базовая матрица штрафов  $\tilde{\mathbf{F}}$ . Матрица штрафов  $\mathbf{F}$  для построения модели выражается через  $\tilde{\mathbf{F}}$ :

$$\mathbf{F} = \gamma \tilde{\mathbf{F}}, \quad (11)$$

где  $\gamma \geq 0$  — весовой множитель, регулирующий допустимую степень несоответствия построенной кластеризации и экспертной.

Базовая матрица штрафов  $\tilde{\mathbf{F}}$

Из \ В	(+, +)	(+, —)	(—, —)
(+, +)	$\delta_{11} = 0$	$\delta_{12} = 0,002$	$\delta_{13} = 0,005$
(+, —)	$\delta_{21} = -0,001$	$\delta_{22} = 0$	$\delta_{23} = 0,003$
(—, —)	$\delta_{31} = -0,003$	$\delta_{32} = -0,002$	$\delta_{33} = 0$

Значения элементов  $\delta$  в матрице  $\tilde{\mathbf{F}}$  определяли исходя из следующих эвристических правил.

1. Переносы, не изменяющие кластерную структуру (например  $(+, -) \rightarrow (+, -)$ ), не штрафуются, потому что они не влияют на отличие алгоритмической модели от экспертной.

2. Чем больше отличий от экспертной кластеризации вносит перенос документа, тем больше порог для осуществления переноса.

3. Должно выполняться свойство транзитивности. Иными словами порог у действия  $(-, -) \rightarrow (-, +)$ ,  $(-, +) \rightarrow (+, +)$  должен совпадать с порогом действия  $(-, -) \rightarrow (+, +)$ .

4. Переносы, возвращающие документы в экспертные кластеры, поощряются.

5. Чтобы избежать циклов, сумма порогов для циклических переносов документа вида  $(+, +) \rightarrow (+, -)$ ,  $(+, -) \rightarrow (+, +)$  должна быть больше нуля.

## 6. Вычислительный эксперимент

Для проверки работы предложенных алгоритмов проводили верификацию тематической модели конференции EURO 2013. В качестве исходных данных был взят набор из 2313 тезисов данной

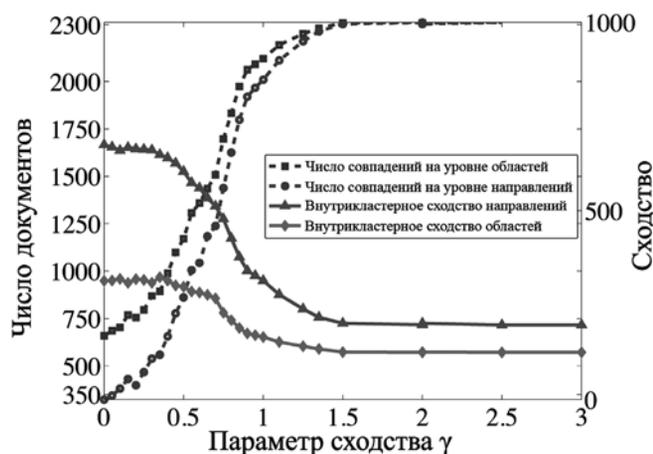


Рис. 3. Зависимость внутри- и межкластерного сходства на уровнях областей и направлений от параметра штрафа  $\gamma$

конференции и модель, построенная экспертами. Модель включала четыре уровня иерархии:  $h = 4$ .

Коллекция  $D$  кластеризовалась алгоритмом, описанным в разд. 5 с параметром  $\alpha = 0,1$  оптимизируемой функции  $Q$  (9). Результаты кластеризации, соответствующие матрице штрафов  $F$  (11) для разных значений  $\gamma$  приведены на рис. 3. По левой оси отложено число документов, для которых экспертная и алгоритмическая кластеризации совпали, по правой оси — значения среднего внутрикластерного сходства (4), а по нижней оси — соответствующее значение параметра  $\gamma$ . Чем больше  $\gamma$ , тем меньше документов попадают в чужие кластеры, но и внутрикластерное сходство становится меньше. Так, при значении  $\gamma > 2$ , 99 % документов попадают в свои экспертные кластеры.

Приведем графики для значения параметра штрафа  $\gamma = 0,7$ , иллюстрирующие изменение сходства между всеми кластерами для уровня областей (рис. 4, см. третью сторону обложки), и распределение документов из экспертных кластеров по алгоритмическим (рис. 5, см. четвертую сторону обложки).

На рис. 4 по осям отложены номера областей. Цвет клетки с координатами  $(x, y)$  соответствует значению среднего сходства между документами области с номером  $x$  и области с номером  $y$ . Клетки диагонали  $(x, x)$  соответствуют внутрикластерному сходству, а клетки  $(x, y)$ ,  $x \neq y$  — межкластерному. Из графиков видно, что внутрикластерное сходство увеличилось в 2 раза (см. рис. 3) при  $\gamma = 0,7$  и при  $\gamma = 2,5$ . При этом увеличение внутрикластерного сходства происходит для большинства кластеров (см. рис. 4).

На рис. 5 клетка с координатами  $(x, y)$  показывает число документов, которые эксперт отнес к кластеру с номером  $x$ , а алгоритм — к кластеру

с номером  $y$ . Большие значения на диагонали  $x = y$  и маленькие вне диагонали показывают, что 67 % документов остались в исходных (экспертных) кластерах. Таким образом, построенная модель схожа с экспертной.

## Заключение

В данной работе предлагался метод анализа и верификации экспертной тематической модели крупной конференции. Были предложены метод составления терминологического словаря конференции и метод построения алгоритмической иерархической тематической модели с учетом экспертной тематической модели. Работа предложенных методов продемонстрирована верификацией экспертной тематической модели конференции EURO 2013.

Работа поддержана проектом РФФИ, грант № 14-07-31264.

## Список литературы

- Hofmann T. Probabilistic latent semantic indexing // Proc. of the 22nd annual international ACM SIGIR conference on research and development in information retrieval. New York: ACM, 1999. P. 50—57.
- Кузьмин А. А., Адуенко А. А., Стрижов В. В. Выбор признаков и оптимизация метрики при кластеризации коллекции документов // Известия Тульского государственного университета. Естественные науки, 2012. № 3. С. 119—131.
- Hartigan J. A., Wong M. A. Algorithm as 136: A k-means clustering algorithm // Applied statistics. 1978. Vol. 28. P. 100—108.
- Загоруйко Н. Г., Ёлкина В. Н., Лбов Г. С. Алгоритмы обнаружения эмпирических закономерностей. Новосибирск: Наука, 1985.
- Pal N. R., Bezdek J. C. On cluster validity for the fuzzy c-means model // IEEE Transactions on Fuzzy Systems. 1995. Vol. 3 (3). P. 370—379.
- Загоруйко Н. Г. Прикладные методы анализа данных и знаний. Новосибирск: Издательство И. М., 1999.
- Борисова И. А. Использование fris-функции для построения решающего правила и выбора признаков (задача комбинированного типа dx) // Труды Всероссийской конференции "Знания—Онтологии—Теории". Новосибирск: Институт математики им. Соболева СО РАН, 2007. Т. 1. С. 37—44.
- Tibshirani R., Hastie T. Discriminative adaptive nearest neighbor classification // IEEE transactions on pattern analysis and machine intelligence. June 1996. Vol. 18, Is. 6. P. 607—616.
- Peng J., Gunopulos D., Domenciconi C. An adaptive metric machine for pattern classification // Advances in Neural Information Processing Systems 13. MIT Press, 2000. P. 458—464.
- Zagoruiko N. G. Methods of recognition based on the function of rival similarity // Pattern recognition and image analysis. 2008. Vol. 18 (1). P. 1—6.
- Daud A., Li J., Zhou L., Muhammad F. Knowledge discovery through directed probabilistic topic models // Frontiers of computer science in China. 2010. Vol. 4 (2). P. 280—301.
- Blei D. M., Ng A. Y., Jordan M. I. Latent dirichlet allocation // Journal of Machine Learning Research. 2003. Vol. 3. P. 993—1022.
- Кузьмин А. А., Стрижов В. В. Проверка адекватности тематических моделей коллекции документов // Программная инженерия. 2013. № 4. С. 16—20.
- EURO 2013: тезисы конференции. URL: <http://euro2013.org/wp-content/uploads/program-euro26.pdf>, дата обращения: 26.12.2013.

УДК 004.315

**С. Г. Бобков**, д-р техн. наук, зав. отделением, e-mail: s\_g\_bob@mail.ru,  
**С. И. Аряшев**, канд. техн. наук, зав. отделом, **М. Е. Барских**, зав. сектором,  
**П. С. Зубковский**, зав. сектором, **Е. В. Ивасюк**, науч. сотр.,  
НИИСИ РАН, г. Москва

## Высокопроизводительные расширения архитектуры универсальных микропроцессоров для ускорения инженерных расчетов

*Компьютерное моделирование многомасштабных процессов горения и детонации требует суперЭВМ эксафлопсной производительности. В работе рассматривается современный подход к моделированию процессов горения, приводится архитектура высокопроизводительных арифметических расширений фирмы Intel и разрабатываемых в НИИСИ РАН, а также предлагается набор решений для создания специализированного потокового сопроцессора, реализуемого в базе самосинхронной логики.*

**Ключевые слова:** эксафлопсная ЭВМ, задача горения, расширение архитектуры микропроцессора, потоковый сопроцессор, самосинхронная логика

S. G. Bobkov, S. I. Aryashev, M. E. Barskyh, P.S. Zubkovskiy, E. V. Ivasyuk

## High-Performance Extensions of Microprocessor Architecture for Speeding-Up of Scientific and Engineering Calculations

*Computer simulation of multiscale burning and detonation processes requires an exaflop-scale performance supercomputer. The paper describes state-of-the-art approach for burning process simulation, architectures of high-performance DSP extensions from Intel and SRISA RAS. Also a number of solutions for dataflow co-processor development based on self-synchronous circuits is proposed.*

**Keywords:** exa-scale computing, burning task, microprocessors architecture extensions, dataflow co-processor, self-synchronous circuits

### Введение

В настоящее время развитие ядерной и термоядерной энергетики, электроники, авиа- и ракетостроения, биотехнологий стало невозможным без проведения полномасштабных инженерных расчетов таких сложных технических и биологических систем с учетом атомно-молекулярного взаимодействия, требующих уже в среднесрочной перспективе (до 2020 г.) использования суперЭВМ эксафлопсного класса ( $10^{18}$  оп/с). Ведущими государствами мира ведутся интенсивные работы по созданию эксафлопсных технологий, а создание суперЭВМ является одним из актуальнейших направлений развития техники. Существующие универсальные микропроцессоры не позволяют достичь требуемой производительности. Одним из методов повышения производительности микропроцессоров является их ориентация на решаемые задачи, такие как задачи

криптографии, обработки сигналов, математические вычисления и др. Существуют два основных направления создания высокопроизводительных микропроцессоров — это создание многоядерных микропроцессоров со встроенным графическим процессором и использование специализированных сопроцессоров, ориентированных на выделенные математические функции. К первому направлению можно отнести проект компании AMD с кодовым именем Seattle (ARM-ядра + графический процессор); европейский проект Mont-Blanc создания эксафлопсной суперЭВМ на базе архитектуры ARM и графических процессоров; микропроцессор Sandy Bridge компании Intel. Архитектура универсальных микропроцессоров со встроенными сопроцессорами широко используется в процессорах многих компаний, одна из подобных архитектур компании Intel рассмотрена ниже. В настоящей работе предла-

гается архитектура процессора, ориентированного на решение задачи горения на базе универсальных RISC-ядер и сопроцессоров, выполняющих основные математические функции задачи горения. Предложена реализация выделенных функций как конвейерное вычисление полиномов.

### **Многомасштабность задачи горения**

Одна из ключевых проблем, требующая решения, — моделирование процессов сгорания новых видов топлива в двигателях новых конструкций. Учет этих процессов требует предсказательного многомасштабного моделирования, включая течение реакции на уровне отдельных молекул и режимы детонации в камере сгорания. Именно многомасштабность процессов горения не позволяет провести их прямое численное моделирование с использованием существующих суперЭВМ и требует достижения производительности экзафлопсного уровня.

В системе моделирования должны одновременно работать совместно различные вычислительные алгоритмы: квантовая молекулярная динамика; классическая молекулярная динамика; кинетический метод Монте-Карло; имитационное моделирование; моделирование больших вихрей; методы с усреднением числа Рейнольдса и др., что в свою очередь требует использования суперЭВМ экзафлопсного класса.

*Первым характерным масштабом* явления в целом при горении смеси горючего и окислителя в камерах сгорания или природных условиях является конвективный перенос исходных реагентов и продуктов реакции в пределах расчетной области. Для двигателей этот масштаб измеряется метрами, а для природных условий — километрами, т. е. соответствует масштабу нефтяного месторождения или области природной катастрофы [1, 2].

*Второй характерный масштаб* обусловлен присутствием турбулентности в потоках и определяется масштабом турбулентных пульсаций (от сантиметров в технических устройствах до десятков метров в атмосфере, характерный масштаб неоднородностей в подземной гидродинамике при реализации внутрипластового горения укладывается в этот диапазон).

*Третий характерный масштаб* обусловлен многофазностью процесса, когда горючее и окислитель находятся в разных фазах (например, жидкое или твердое горючее в форме капель или частиц и газообразный окислитель, или же газообразное горючее и жидкий окислитель в случае водородного двигателя). Этот масштаб характеризуется размером гетерогенных включений: капель, частиц или горючих фрагментов (от долей миллиметров в технических устройствах до метров в природных условиях). При распространении горения в пористой среде возникает дополнительный масштаб, связанный с характерным размером пор.

*Четвертый характерный масштаб* определяется процессами молекулярной диффузии реагентов и теплоты и связан с толщиной зоны пламени (от миллиметров до сантиметров).

*Пятый характерный масштаб* связан с собственно химическими взаимодействиями между молекулами реагентов (нанометры), и определяет скорость химической реакции, влияние промоторов, ингибиторов, адсорбции и десорбции в случае гетерогенных реакций. При прямом численном моделировании таких процессов используют методы молекулярной динамики, в которых ключевым моментом является правильное задание потенциалов взаимодействия между молекулами различных веществ.

Для расчета потенциалов взаимодействия необходимо спуститься на уровень меньших масштабов и проводить расчеты методами квантовой механики на уровне одной молекулы. Это шестой характерный масштаб данной задачи.

Следует отметить, что прямое увеличение разрешающей способности кода за счет увеличения числа расчетных узлов по всему полю задачи неэффективно. Даже при наличии супервычислителя, прямое уменьшение расчетных ячеек приведет к соответствующему увеличению числа шагов по времени, что при долгосрочном прогнозировании вызовет накопление ошибки вычислений, сводящей к нулю преимущества супервычислителя. Эксафлопсная производительность требуется для сопряжения расчета многомасштабных процессов с заданной точностью в темпе собственного времени каждого процесса, применения наиболее эффективных алгоритмов для решения разных классов задач, обеспечения эффективного обмена данными, а не только для получения возможности прямого увеличения количества вычислительных операций.

В целом наблюдается общая тенденция смещения акцентов при моделировании процессов горения либо в сторону более точного учета кинетики химических реакций, либо в сторону более детального моделирования газодинамических процессов. Предполагаемая система экзафлопсного класса, с архитектурой и системой обменов, ориентированных на решение конкретных задач горения, позволит проводить сбалансированное моделирование многомасштабных процессов с учетом кинетических газодинамических и межфазных процессов с одинаковой точностью.

### **Профилирование задачи горения**

В настоящее время компьютерное моделирование процессов горения и детонации выполняется путем распараллеливания вычислительных программ на многопроцессорных ЭВМ и графических ускорителях [3]. С помощью профилирования можно получить информацию о наиболее часто используемых в таких задачах функциях или функциях, вычисление которых занимает наибольшее

Таблица 1

## Результаты профилирования задачи горения

Время выполнения, %	Название функции	Примечание
20,93	GetTransport	Коэффициенты переноса для газовой смеси на основании молярных масс или плотностей компонентов
11,67	__ieee754_exp	Библиотечная функция вычисления вещественной экспоненты (библиотека libm-2.17.so)
6,88	UpdateExtrate	Добавка данных от элементарной реакции в интенсивность образованного компонента
6,70	ElemCofs	Элементарный механизм: коэффициенты для прямой и обратной реакции
5,11	GetThermo	Явная стадия турбулентности Вычисление якобиана
4,02	ExplTurbStep	
3,53	GetJacobean	
3,47	__pow_finite	Библиотечная функция возведения в степень (библиотека libm-2.17.so)
2,59	GetDifComp	Учет диффузии компонентов
2,53	Y2X	Вычисление плотности компонентов в каждой ячейке

время. В табл. 1 приведены результаты профилирования программ метода ETDRK4 для системы ОДУ с кинетикой Мааса и Варнаца (горение водорода, 9 компонентов, 19 реакций) и газодинамический код с учетом той же кинетики, но решение системы осуществляется четырехстадийным методом Роземброка. Запуск выполняли на ЭВМ архитектуры x86 с процессором Intel(R) Core(TM)2 Duo CPU E6550 2.33GHz. В первом столбце табл. 1 указано время выполнения функции в процентах от общего времени выполнения программы, во втором столбце — название функции, а в третьем — примечания.

Анализ кода функций, наиболее часто применяемых в рассматриваемой реализации задачи горения, показывает, что в большинстве случаев выполняются операции умножения и сложения над матрицами и векторами, векторные операции деления и извлечения квадратного корня. Вызовы функций вычисления экспоненты и возведения в степень происходят в цикле над отдельными элементами вектора, с использованием, таким образом, только скалярных функций.

Набор функций, описанный выше на частном примере реализации задачи горения, является характерным для многих вычислительных задач, и именно время выполнения этих операций наряду с возможностями подсистемы памяти характеризует производительность микропроцессора при инженерных и научных расчетах.

### Возможности современных специализированных и универсальных микропроцессоров

Область применения современных специализированных микропроцессоров, ориентированных на

цифровую обработку сигналов (DSP) ограничена обработкой аудио- и видеопотоков в мобильных и персональных коммерческих устройствах. Примером таких микропроцессоров может служить семейство Xtensa компании Tensilica. В микропроцессорах Xtensa реализована 32-разрядная RISC-архитектура с пятистадийным конвейером и 16/24-битными полями инструкций. Вычислительные возможности этих микропроцессоров представлены целочисленными 16- и 32-разрядными умножителями и сумматорами, 16-разрядным блоком умножения с накоплением и блоком вещественной арифметики одинарной точности. Эти возможности отвечают современным требованиям по обработке аудио- и видеоданных в персональных устройствах, но не являются достаточными для применения микропроцессоров семейства Xtensa для высокопроизводительных научных и инженерных вычислений.

Более широкими вычислительными возможностями обладают DSP микропроцессоры семейства TigerShark компании Analog Devices. Микропроцессор ADSP-TS201S при тактовой частоте 600 МГц демонстрирует производительность 3,6 Гфлопс на вещественных операциях одинарной точности и может выполнять до двух 32-разрядных операций умножения с 80-разрядным накоплением за такт. Отсутствие вещественных операций двойной точности существенно ограничивает применение таких микропроцессоров для высокопроизводительных вычислений, где двойная точность является базовым требованием, а в ряде случаев используется удвоенная двойная точность (*double-double*) и четверная точность (*quad precision*).

Высокопроизводительные коммуникационные микропроцессоры NetLogic XLP основаны на архитектуре MIPS64. Наиболее мощный микропроцессор XLP980 выполнен по проектным нормам 28 нм, содержит 20 процессорных ядер, работающих на частоте 2 ГГц. Четыре канала Ethernet 40 Гбит обладают пропускной способностью 160 Гбит/с, а внутренняя накристалльная сеть позволяет объединить до восьми ядер в кластер с когерентными кэшами и связями между ядрами со скоростью 1,28 Тбит/с.

В большинстве систем для проведения высокопроизводительных научных и инженерных расчетов в настоящее время не используются специализированные микропроцессоры для повышения производительности, а строятся на основе универсальных микропроцессоров и параллельных многопоточковых вычислителей, таких как графические ускорители или векторные сопроцессоры. В гибридных системах используются микропроцессоры фирм Intel и AMD и графические ускорители Nvidia и AMD.

Среди векторных сопроцессоров стоит отметить сопроцессор NEON, разработанный фирмой ARM, и развитие архитектуры AltiVec — сопроцессор VSX микропроцессора IBM Power7.

Более подробно первый подход с использованием векторных расширений стандартной архитектуры может быть рассмотрен на примере разработок фирмы Intel для архитектуры x86.

### **Векторные расширения в микропроцессорах фирмы Intel**

В микропроцессорах фирмы Intel векторные расширения начали применять с 1997 г., когда в процессоре Pentium было представлено расширение MMX, предназначенное для ускорения обработки изображений и воспроизведения потокового аудио и видео. Набор команд MMX включает в себя целочисленные арифметические, логические команды, команды сравнения, сдвига, преобразования данных [4]. В расширении MMX использовано восемь 64-разрядных регистра MM0...MM7, отображенных на регистровый файл FPU. Команды MMX работали с 64-разрядными целочисленными данными либо с данными, упакованными в вектор шириной 64 разряда. Это могли быть 32-разрядные слова, упакованные по 2, 16-разрядные слова, упакованные по 4, или 8 байт. Следующим векторным расширением архитектуры x86 стала архитектура SSE, представленная в процессоре Pentium II. Расширение SSE было призвано решить две главные проблемы, свойственные MMX: невозможность использования FPU одновременно с MMX и поддержка только целочисленных типов данных. Архитектура SSE включает в себя восемь самостоятельных 128-разрядных регистров, в которых содержатся вещественные данные одинарной точности. Данные могут быть скалярными или упакованными в вектор. В наборе инструкций появились 70 новых команд для данных с плавающей запятой, а также 128-разрядные команды обмена с памятью. Архитектура SSE2 объединила возможности MMX и SSE, дополнив их операциями над вещественными числами двойной точности. 128-разрядные регистры XMM0...XMM7 могут содержать два последовательных числа с плавающей запятой двойной точности.

Расширение AVX построено в соответствии с архитектурой SIMD и позволяет выполнять арифметические операции с плавающей запятой одинарной и двойной точности над векторами изменяемой ширины до 256 разрядов с использованием данных из регистровых файлов или из локальной памяти. В AVX, в отличие от предыдущих расширений архитектуры x86, появилась возможность выполнять неразрушающие арифметические команды, т. е. команды с тремя операндами.

### **Микропроцессоры Xeon Phi компании Intel**

В 2013 г. фирмой Intel была представлена архитектура сопроцессора Intel® Xeon Phi™, позволившая совершить большой скачок в производительности и функциональности векторных вычисле-

ний. Базовой архитектурой для сопроцессора Xeon Phi™ является архитектура Intel® Many Integrated Core (Intel® MIC), а вычислительным ядром является векторный исполнительный блок (VPU — *vector processing unit*). Ширина обрабатываемого вектора составляет 512 бит, за один такт могут быть выполнены 16 вещественных операций одинарной точности или 8 вещественных операций двойной точности. Помимо этого, в архитектуре Intel® MIC предусмотрены команды FMA (*fused multiply-add*), выполняющие 32 вещественные операции одинарной точности или 16 вещественных операций двойной точности. Также предусмотрена поддержка целочисленных операций и вычисления трансцендентных функций, таких как обратная величина, обратный квадратный корень, логарифм и экспонента с основанием 2 [5].

Наибольшую эффективность архитектура Intel® Xeon Phi™ демонстрирует на задачах высокопроизводительных вычислений (HPC — *high performance computing*), запуск же отдельных инструкций не оправдывает затраты на выборку, декодирование и выполнения других вспомогательных стадий конвейера. Для повышения эффективности работы с широкими векторами в архитектуре Intel® Xeon Phi™ предусмотрена аппаратная реализация таких функций, как наложение маски на обрабатываемые векторы, нелокализованные обращения к памяти для сборки или распределения вектора по памяти и облегчения векторизации программного кода.

Пиковая производительность микропроцессора Intel® Xeon Phi™ семейства 7100, содержащего 61 ядро и имеющего тактовую частоту 1,278 ГГц, составляет 1,2 Тфлопс для операций двойной точности (16 операций X 61 ядро X 1,278 ГГц). Данные производителя о реальной производительности на синтетических тестах перемножения матриц DGEMM — 1,05 Тфлопс [6].

Реализация двумерного преобразования Фурье двойной точности на процессоре Intel® Xeon Phi™ с использованием математической библиотеки Intel® MKL демонстрирует производительность 103 Гфлопс [7].

Тесты производительности, основанные на вычислении полиномов по схеме Горнера, показывают существенную зависимость эффективности вычислений от степени полинома и от числа исполнительных потоков на ядро. Для полинома 10-й степени удается достичь максимум 50 % эффективности, в то время как увеличение степени полинома до 32 повышает эффективность до 70 % от пиковой. Максимальные значения были получены при распараллеливании четырех потоков на ядро [8].

Отличительной особенностью вычисления экспоненты и логарифма в сопроцессоре Intel® Xeon Phi™ является ограничение по точности получаемого результата. Вычисление экспоненты проводится по основанию 2 для целых 24-разрядных чисел с

точностью, не превышающей 0,99 ulp (младший значащий бит). Логарифм по основанию 2 вычисляется для вещественных чисел одинарной точности, точность вычислений составляет  $2^{-21}$  в диапазоне входного операнда [0,5, 2] и 4 ulp вне этого диапазона.

### Сопроцессоры микропроцессоров серии КОМДИВ

Из российских разработок можно выделить семейство высокопроизводительных микропроцессоров серии 1890 (КОМДИВ), работа над которыми ведется в НИИСИ РАН. В микропроцессорах 1890BM6Я, 1890BM7Я, а также в перспективной разработке 1890BM8 применены различные архитектурные решения для повышения производительности при проведении научных и инженерных расчетов.

Сопроцессор CP2 процессора K128РИО представляет собой SIMD-устройство, ориентированное на выполнение задач цифровой обработки сигналов (рис. 1). Пиковая производительность сопроцессора составляет 8 Гфлопс одинарной точности для тактовой частоты 200 МГц. Сопроцессор содержит четыре вычислительные секции, каждая из которых может выполнять до 10 вещественных операций одинарной точности за такт на операции комплексного умножения с прибавлением и вычитанием третьего аргумента, более известной как "бабочка Фурье". В сопроцессоре предусмотрена внутренняя память инструкций и коэффициентов, а также память данных в каждой вычислительной секции. Каждая вычислительная секция содержит скалярный блок специальных функций, позволяющий выполнять операции вычислений обратной величины, обратного квадратного корня, логарифма, экспоненты, тригонометрических функций.

Сопроцессор комплексной арифметики CP2 процессора K64РИО [9, 10] представляет собой специализированный сопроцессор, ориентированный на выполнение арифметических операций над комплексными и вещественными числами с плавающей запятой одинарной точности (рис. 2). Сопроцессор выполнен по архитектуре SIMD2, т. е. имеет две вычислительные секции, использует стандартный конвейер выдачи команд и может выполнять одну арифметическую команду и одну команду обмена данными с памятью или ядром процессора над двумя потоками данных. Вычислительное ядро сопроцессора позволяет выполнять до 10 операций над

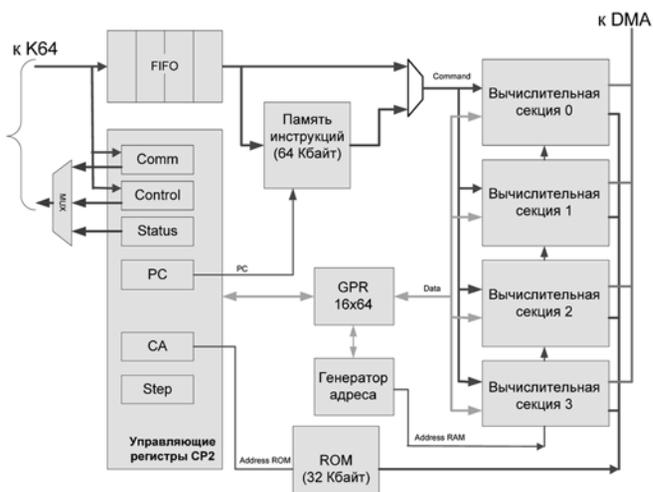


Рис. 1. Структурная схема CP2 процессора K128РИО

вещественными числами одинарной точности за такт на каждой секции. Таким образом, пиковая производительность сопроцессора на частоте 250 МГц составляет 5 Гфлопс одинарной точности. Отличительной особенностью сопроцессора является использование блоков FMA в вычислительном ядре.

Развитием сопроцессора комплексных вычислений стал векторный сопроцессор в процессоре K64M (рис. 3). Вычислительный блок векторного сопроцессора позволяет выполнять операции одинарной и двойной точности над векторами шириной 128 разрядов. Пиковая производительность достигается при использовании команд комплексного умножения с прибавлением и вычитанием третьего операнда ("бабочка Фурье") и составляет за такт 10 вещественных операций двойной точности или 20 вещественных операций одинарной точности.

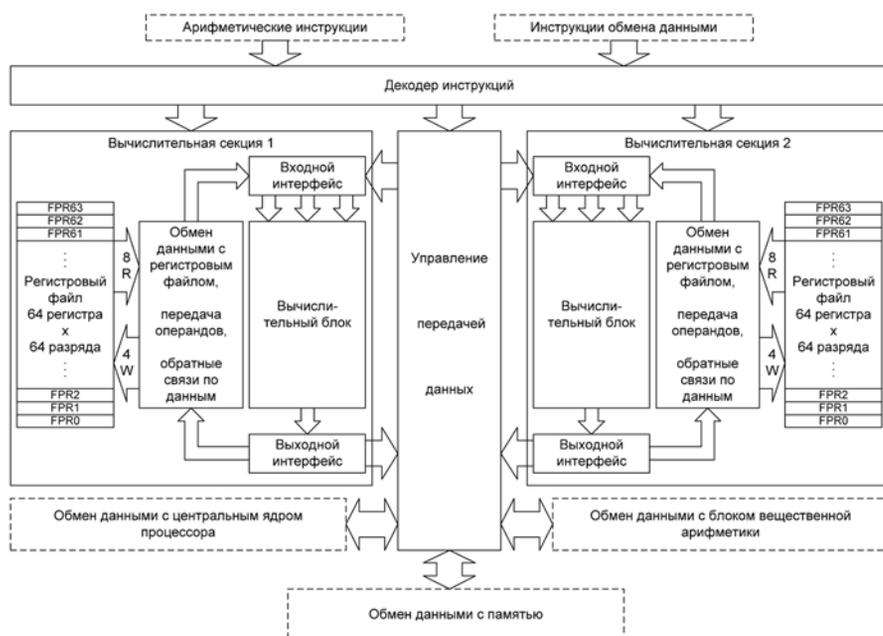


Рис. 2. Структурная схема сопроцессора процессора K64РИО



Рис. 3. Структурная схема векторного сопроцессора процессора K64M

венных операций одинарной точности. Отказ от использования локальной специализированной памяти данных (как в сопроцессорах K128PIO и K64PIO) позволяет использовать вычислительные возможности сопроцессора на пользовательских приложениях, скомпилированных стандартными средствами с подключением математических библиотек.

### Оптимизация библиотеки BLAS для векторного сопроцессора

Важным условием для эффективного использования векторного сопроцессора на пользовательских задачах является наличие математической библиотеки, оптимизированной под архитектуру сопроцессора. Ряд функций из широко используемой библиотеки линейной алгебры BLAS были оптимизированы под архитектуру сопроцессора.

Таблица 2

Результаты измерения эффективности выполнения функции GEMM векторным сопроцессором, %

$M \times N \times K$	L2	ZGEMM	CGEMM	DGEMM	SGEMM
4 × 4 × 8	0,34	23,92	н/д	н/д	н/д
8 × 4 × 8	0,59	34,12	н/д	н/д	н/д
32 × 4 × 8	2,05	52,24	н/д	н/д	н/д
256 × 4 × 8	15,72	60,72	н/д	н/д	н/д
4 × 8 × 8	0,6	30,93	21,18	14,21	н/д
8 × 8 × 8	1,0	40,59	31,07	22,38	н/д
32 × 8 × 8	3,32	55,05	47,19	42,16	н/д
256 × 8 × 8	25,2	61,08	55,52	57,72	н/д
4 × 8 × 16	0,88	35,88	25,83	17,69	11,37
8 × 8 × 16	1,37	45,13	35,65	26,99	18,73
32 × 8 × 16	4,3	57,95	49,91	46,34	38,04
256 × 8 × 16	31,64	62,79	57,15	58,76	55,12
16 × 16 × 16	3,9	55,30	47,44	42,24	33,33
1024 × 4 × 8	62,6	61,73	н/д	н/д	н/д
1024 × 8 × 8	100,2	61,83	56,56	60,20	н/д
4096 × 8 × 8	400,2	37,48	н/д	42,41	н/д
256 × 32 × 32	103,12	63,92	57,95	61,78	58,47
64 × 64 × 64	62,5	63,11	56,61	59,02	54,73
256 × 64 × 64	212,5	62,15	58,11	61,96	59,53
1024 × 8 × 16	125,39	н/д	58,06	61,66	57,98
4096 × 8 × 16	500,4	н/д	н/д	38,22	н/д
1024 × 16 × 16	200,8	н/д	58,15	61,86	58,29

Результаты профилирования тестов производительности LinPack показывают, что наиболее используемыми функциями библиотеки BLAS (более 90 % от времени исполнения) являются функции DGEMM — перемножение матриц, DTRSM — решение системы линейных дифференциальных уравнений и DGETRF — LU-разложение, причем функция DGEMM является базовой для DTRSM и DGETRF при использовании блочных методов.

В табл. 2 приведены результаты измерения эффективности использования векторного сопроцессора для функции GEMM, реализованной на языке Ассемблер с использованием команд векторного сопроцессора.

В первом столбце приведены размеры перемножаемых матриц, во втором — степень заполнения кэш-памяти второго уровня. Следующие столбцы содержат результаты для разных типов входных данных — комплексных двойной точности, комплексных одинарной точности, вещественных двойной точности, вещественных одинарной точности. Эффективность использования векторного сопроцессора оценивали по формуле  $Eff = N_{RTL}/N$ , где  $Eff$  — искомая эффективность;  $N_{RTL}$  — число машинных тактов, затраченных на выполнение теста с вызовом функции;  $N$  — теоретически вычисленное число арифметических команд сопроцессора, выдаваемых в каждом такте. Максимальное значение эффективности приближается к 64 % для случая, когда исходные данные полностью помещаются в кэш-память второго уровня, число арифметических команд достаточно велико, чтобы скрыть затраты на передачу данных, а комплексный тип данных позволяет использовать наиболее производительные команды сопроцессора.

Для достижения высокой производительности при проведении инженерных расчетов необходимо не только наличие оптимизированной под конкретную архитектуру математической библиотеки линейной алгебры, но и возможность аппаратного ускорения вычисления трансцендентных функций. Выполнение таких расчетов в целочисленном формате или в формате с плавающей запятой одинарной точности уже не отвечает постоянно возрастающим требованиям к точности инженерных расчетов.

### Оценки погрешности при аппаратной реализации аппроксимирующего полинома

В условиях ограниченной разрядности представления данных в ЭВМ, трансцендентные функции не могут быть вычислены точно, а могут быть только аппроксимированы. Одним из распространенных алгоритмов вычисления упомянутых функций является аппроксимация полиномом/рациональной дробью или разложение в степенной ряд. Ключевым параметром в проектировании алгоритма аппроксимации является максимальная допусти-

Число бит мантиссы, соответствующее абсолютной ошибке аппроксимации (минимаксный алгоритм подбора коэффициентов) в диапазоне входного аргумента [0,1]

Функция	Степень полинома							
	2	3	4	5	6	7	8	9
sin(x)	7,8	12,7	16,1	21,6	25,5	31,3	35,7	41,9
$e^x$	6,8	10,8	15,1	19,8	24,6	29,6	34,7	40,1
$\ln(1+x)$	8,2	11,1	14,0	16,8	19,6	22,3	25,0	27,7
$(x+1)^x$	6,3	8,5	11,9	14,4	18,1	20,0	22,7	25,1
arctan(x)	8,7	9,8	13,2	15,5	17,2	21,2	22,3	24,5
tan(x)	4,8	6,9	8,9	10,9	12,9	14,9	16,9	19,0
$x^{1/2}$	3,9	4,4	4,8	5,2	5,4	5,6	5,8	6,0
arcsin(x)	3,4	4,0	4,4	4,7	4,9	5,1	5,3	5,5

мая ошибка вычисления. От этой ошибки в первую очередь зависит сложность реализации алгоритма, какой бы он ни был: программный, программно-аппаратный или полностью аппаратный.

Так как диапазон изменения аргумента, от которого требуется найти значение функции, имеет большую протяженность (согласно стандарту IEEE754), то для обеспечения допустимого значения погрешности на всем диапазоне требуется обеспечить очень большую степень аппроксимирующего полинома или большое число членов ряда разложения. Это потребует больших временных затрат на вычисление и поэтому для упрощения полиномов/рядов применяют сужение диапазона входного аргумента, называемое приведением аргумента (*range reduction, argument reduction*) [11]. Затем осуществляется вычисление полинома от приведенного аргумента с заданной точностью и восстановление результата до соответствующего изначальному диапазону изменения аргумента.

В табл. 3 приведено число "точных" битов мантиссы результата аппроксимации, соответствующее абсолютной ошибке приближения (полученное как  $\log_2(abs\_err)$ ) для основных элементарных функций, при различных степенях аппроксимирующего многочлена для диапазона входного аргумента [0,1] [11].

Из табл. 3 видно, что даже для приведенного аргумента ([0,1]), чтобы обеспечить получение результата в формате *double* (IEEE754,  $2^{-52} \approx 10^{-16}$ ) с погрешностью 1ulp и более, необходим полином степени выше 10.

Показательная функция  $e^x$  является самой простой с точки зрения вычисления на ЭВМ, так как полином, обеспечивающий ее аппроксимацию, демонстрирует большую скорость сходимости при увеличении степени полинома (рис. 4), а также диапазон входного аргумента, при котором функция не вызывает исключений, предусмотренных стандартом IEEE754, узок относительно полного диапазона стандарта.

Так, для формата *double* диапазон, в котором  $e^x$  не вызывает переполнения и потери значимости, составляет:

1.  $[X_{lower}, X_{upper}] = [\ln(X_{min}), \ln(X_{max})] = [-708, +709]$ , если не обрабатываются денормализованные числа.

2.  $[X_{lower\_den}, X_{upper}] = [\ln(X_{min\_den}), \ln(X_{max})] = [-745, +709]$ , если обрабатываются денормализованные числа.

В настоящее время широко применяется полиномиальная аппроксимация функций по минимаксному принципу, в которой используется алгоритм Ремеза расчета коэффициентов. Минимаксная аппроксимация функции обеспечивает минимизацию максимальной ошибки на заданном диапазоне. Расчеты полиномов сильно упрощаются благодаря использованию таких программ компьютерной алгебры, как, например, Maple.

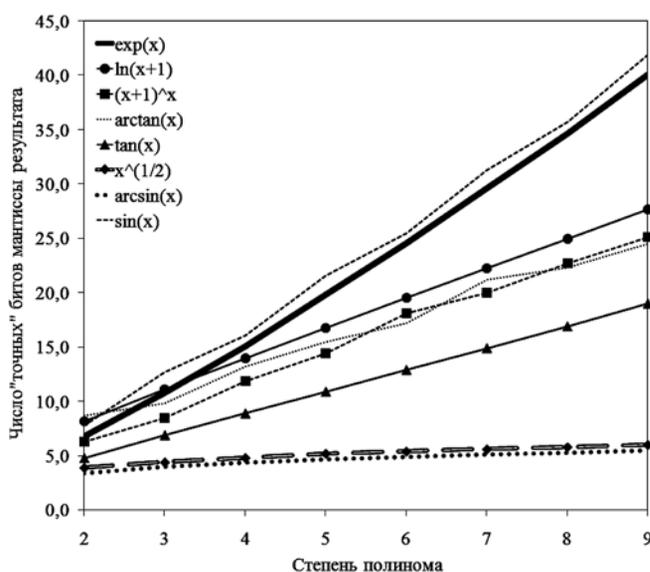


Рис. 4. Зависимость числа "точных" битов аппроксимации от степени полинома для различных функций

Как было сказано выше, главным параметром в проектировании алгоритма вычисления специальных функций и  $e^x$  в частности, является максимальная ошибка аппроксимации. Она складывается из погрешностей стадии приведения аргумента, погрешности стадии вычисления полинома и погрешности стадии восстановления/округления. Стандарт IEEE754 в последней редакции (2008 г.) ничего не декларирует об элементарных функциях, но тем не менее, существуют программные библиотеки, обеспечивающие точное округление, например CR-LIBM [12], правильность реализации функции экспоненты в которой доказана формально, и итоговая погрешность вычисления составляет 0,5ulp. Для обеспечения такой ошибки вычисления проводятся с повышенной внутренней разрядностью. Приведем основные параметры реализации функции в CR-LIBM.

1. Результат имеет погрешность не более 0,5ulp. Библиотека ориентирована на микропроцессоры, поддерживающие арифметику типа *double* или

Зависимость числа старших нулей мантиссы приведенного аргумента (Worst Case) и степени аппроксимирующего полинома от значения константы  $C$

$C$ (диапазон аргумента $[-C/2... C/2]$ )	Worst Case, бит	Степень полинома	Ошибка полинома	Разрядность вычислений $r$ , бит (не менее)
$\ln 2 / 32786$	72	2	$4.93e-17$	125
$\ln 2 / 4096$	69	3	$2.67e-19$	122
$\ln 2 / 256$	65	4	$2.37e-18$	118
$\ln 2 / 64$	61	5	$1.1e-18$	114
$\ln 2 / 16$	59	6	$7.09e-18$	112
$\ln 2 / 2$	58	10	$1.23e-19$	111
$\ln 2$	57	11	$4.33e-18$	110
$\ln 2 \cdot 2$	56	14	$3.85e-19$	109
$\ln 2 \cdot 4$	55	17	$1.75e-18$	108
$\ln 2 \cdot 8$	54	22	$2.46e-18$	107
$\ln 2 \cdot 16$	53	30	$4.27e-18$	106

*double-extended*. Внутри эмулируется большая разрядность промежуточного результата: около 120 бит для типичного случая и примерно 160 бит для наихудшего случая, за счет использования арифметики *double-double* и *triple-double*.

2. Применяется сложный двух- или трехступенчатый метод приведения, используется таблица небольшого объема.

3. Диапазон приведенного аргумента таков, что позволяет использовать полином малой степени: 4-го для типичных случаев, не более 7-го для наихудших.

4. Отношение времен выполнения для наихудшего/наилучшего случаев не превышает 7,65.

Для разработки алгоритма  $e^x$  на аппаратном или программно-аппаратном уровне необходимо оценить необходимую разрядность внутреннего представления данных для обеспечения погрешности в  $0,5ulp$  и решить можно ли ее обеспечить на всех стадиях вычисления, используя существующие арифметические модули вещественной арифметики микропроцессора (модуль сложения, модуль умножения). Наиболее часто используемый алгоритм приведения аргумента — алгоритм, основанный на вычитании. Например, в CR-LIBM используется следующее приведение аргумента:

$$r = x - k \cdot C,$$

где  $x$  — входной операнд;  $C = \ln(2)/4096$ ,  $k = \text{int\_rnd}(x \cdot 4096 / \ln(2))$ .

Известно, что основным недостатком алгоритма приведения, основанного на вычитании, является потеря значащих битов при некоторых значениях входного операнда. Близость произведения  $k \cdot C$  к значению входного операнда  $x$  называется наихудшим случаем. Мера этой близости — количество старших нулей результата вычитания. Так, используя программу, предложенную в работе [11], можно оценить *worst case* для  $C = \ln(2)/4096$  и разрядности мантиссы 53 бита:

*The worst case occurs for  $x = 7908852068369537 \cdot 2^{-46}$ ,  
The corresponding reduced argument is:  
0.14617290722938609735451015055194799833960803383  
4331348599962250511382193663584858956059618000  
0385832e-20  
whose radix 2 logarithm is  $-69.21281406$ .*

То есть для числа  $x = 7908852068369537 \cdot 2^{-46}$ , имеющего в формате *double* вид  $405C190F1E1DC481_{HEX}$  число старших нулей при вычитании  $r = x - k \cdot C$  будет составлять 69. Если для подобных вычислений пользоваться числами формата *double*, то все значащие биты приведенной мантиссы будут потеряны. Поэтому для минимизации ошибки приведения, вычисления требуют значительного (двух/трехкратного) увеличения разрядности. А именно, значение внутренней разрядности этапа приведения для обеспечения итоговой ошибки не более  $0,5ulp$

должна составлять  $W_{RED} = N_{ZER\_WORST\_CASE} + W_{MAN\_D} + W_{GUARD}$ , где  $N_{ZER\_WORST\_CASE}$  — число старших нулей при вычитании;  $W_{MAN\_D} = 53$  — разрядность мантиссы формата *double*;  $W_{GUARD}$  — число "охранных" битов, необходимых для округления. Таким образом, для наихудшего случая разрядность внутреннего представления данных на этапе приведения будет составлять как минимум  $W_{RED} = 122$  бита.

В табл. 4 приведены результаты расчета внутренней разрядности в алгоритме приведения аргумента при различных значениях константы. Вычисления проведены с помощью пакета Maple. Ошибка полинома вычислялась при точности коэффициентов 50 десятичных разрядов. Для обеспечения итоговой ошибки вычисления функции не более  $0,5ulp$ , необходимо обеспечить полное сохранение разрядности мантиссы на этапе приведения и поэтому арифметический модуль, выполняющий вычитание, должен обеспечивать необходимое число битов разрядности мантиссы, располагающихся после десятичной точки. Как видно из табл. 4, требуемые разрядности значительно превышают характеристики формата *double*.

#### Особенности использования блоков FMA в схеме вычисления полинома

В векторном сопроцессоре процессора K64M в качестве вычислительного ядра используется блок FMA (Fused Multiply-Add) [10]. Внутренняя разрядность данных в этом блоке составляет 161 бит, что позволяет использовать его для создания схемы вычисления полиномов с точностью  $0,5ulp$ .

Важной особенностью блока FMA является очередность этапов внутренней нормализации и суммирования с распространением переноса. Если в классических модулях умножения с накоплением FMA сначала выполняется суммирование, затем нор-

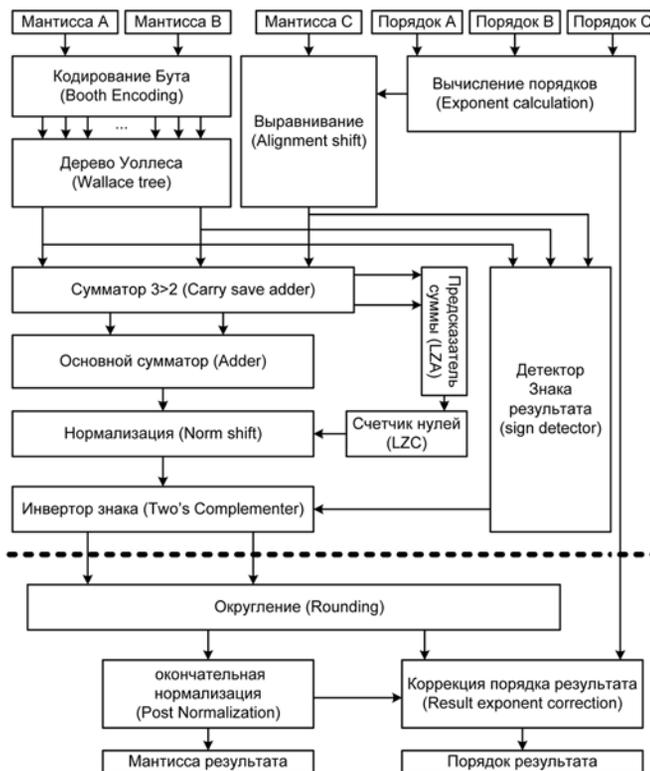


Рис. 5. Структурная схема FMA модуля с нормализацией мантиссы после суммирования

мализация и округление, в описанных ниже модулях сначала выполняется нормализация, а затем финальное суммирование, совмещенное с округлением. Структурные схемы модулей приведены на рис. 5 и рис. 6.

Такая организация стадий обработки имеет потенциал по модернизации модуля, позволяющий эффективнее использовать внутреннюю разрядность мантиссы результата и не терять младшие значащие биты при обработке наихудших случаев приведения аргумента. На рис. 7 приведены иллюстрации, поясняющие внутреннюю организацию положения мантиссы в обычном блоке FMA и в блоке с предложенными изменениями. Из рисунка видно, что в последнем случае, благодаря ранней нормализации, значащие биты промежуточного результата помещаются в левую часть поля представления и остается больше места для результата вычитания при наихудшем случае.

Уменьшить вычисляемую степень полинома или глубину цепочки операций "умножение с накоплением" можно распараллеливанием вычислений. Например, можно привести такой полином к виду, который предполагает одновременное вычисление нескольких его частей:

$$P(x)_{2n} = x^n(((a_{2n}x + a_{2n-1})x + a_{2n-2})x + a_{2n-3}) \dots x + a_{n+1} + (((a_nx + a_{n-1})x + a_{n-2})x + a_{n-3}) \dots x + a_0.$$

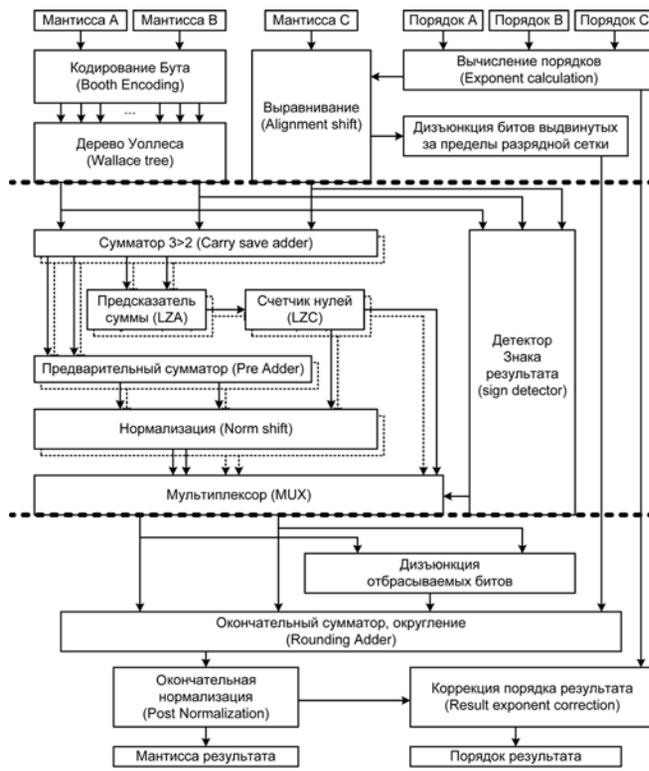


Рис. 6. Структурная схема FMA модуля с нормализацией промежуточных данных до окончательного суммирования

Алгоритм, обеспечивающий вычисление значения функции на основе такого полинома представлен на рис. 7, в. Для реализации такого алгоритма требуется поддержка вычисления функции  $x^N$ .

Вычисление функции можно осуществить также с помощью рациональной аппроксимации, которая позволяет уменьшить расчетную степень аппрок-

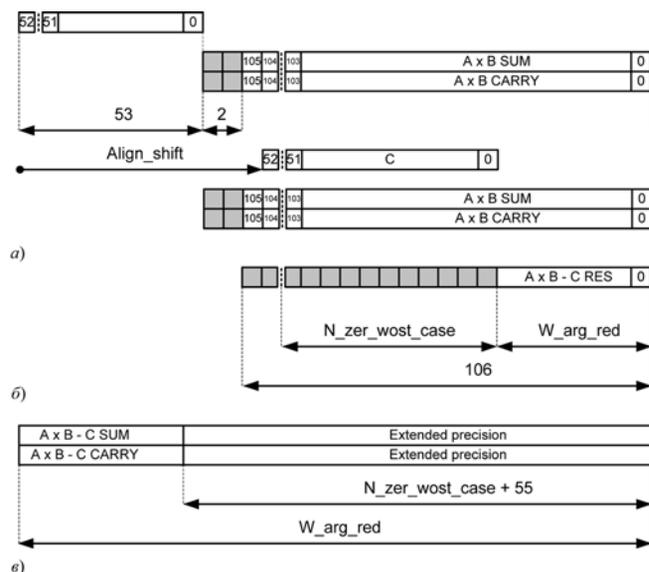


Рис. 7. Внутренняя организация положения мантиссы в модулях FMA до сложения (а), после суммирования и до нормализации в обычном модуле FMA (б), в предложенном модуле с нормализацией до окончательного суммирования (в)

симирующих полиномов числителя и знаменателя примерно вполовину, по сравнению с вариантом, представленным на рис. 7, а. Выражение для вычисления имеет вид

$$P(x)_n = \frac{((\dots((a_n x + a_{n-1})x + a_{n-2})x + a_{n-3})\dots)x + a_0}{((\dots((b_n x + b_{n-1})x + b_{n-2})x + b_{n-3})\dots)x + b_0}$$

Однако требуется поддержка операции деления значений числителя и знаменателя. Алгоритм, обеспечивающий вычисление значения функции на основе рациональной аппроксимации, представлен на рис. 7, б.

Алгоритм, представленный на рис. 7, в, потенциально быстрее алгоритма рис. 7, б, так как операция FMA, завершающая вычисление, существенно быстрее, чем операция DIV, а задержка возведения в степень может быть маскирована задержкой основных конвейеров вычисления полиномов. Структурные схемы, приведенные на рис. 8, полностью конвейерные, задержки вычисления для данных алгоритмов следующие:

Рис. 8, а:

$$T_{total} = T_{red} + N \cdot T_{fma} + T_{reconstr} + T_{rnd}$$

Рис. 8, б:

$$T_{total} = T_{red} + (N \cdot T_{fma})/2 + T_{div} + T_{reconstr} + T_{rnd}$$

Рис. 8, в:

$$T_{total} = T_{red} + (N \cdot T_{fma})/2 + T_{fma} + T_{reconstr} + T_{rnd}$$

где  $T_{red}$  — задержка стадии приведения аргумента,  $T_{fma}$  — задержка модуля FMA;  $T_{div}$  — задержка модуля деления;  $T_{reconstr}$  — задержка стадии восстановления;  $T_{rnd}$  — задержка стадии округления;  $T_{total}$  — общая задержка выполнения алгоритма.

Если в качестве базовой операции этапа приведения будет выбрано "умножение с накоплением", реализуемое модулем FMA, то структурная схема модуля, реализующего весь алгоритм, становится более регулярной (рис. 8, г).

Оценить время вычисления экспоненты предложенными схемами можно для случая, когда диа-

пазон приведения входного аргумента составляет  $[-(\ln 2)/2, (\ln 2)/2]$ . Этот вариант является компромиссным с точки зрения степени аппроксимирующего полинома и разрядности мантиссы в наихудшем случае (см. табл. 4). Время выполнения операции FMA составляет 4 такта, а операции деления — 8 тактов. Для приведения аргумента достаточно использовать одну операцию FMA, а степень аппроксимирующего полинома составляет 11. Точность результата получается не хуже 1ulp. В этом случае время вычисления экспоненты для схемы на рис. 8, а составит примерно 50 тактов, для схемы рис. 8, б — около 35 тактов, а параллельное возведение в степень (рис. 8, в и 8, г) позволяет сократить время выполнения до 30 тактов.

Вычисление значения полинома высокой степени потребует многократного использования блока FMA в конвейере или итерационно, что может внести не меньшую ошибку в вычисления, чем на стадии приведения. Например, значения ошибок вычисления функции, представленные в табл. 4, получены при разрядности коэффициентов 50 десятичных разрядов. Если в качестве "строительного" элемента используется модуль FMA с фиксированной разрядностью входных операндов (*double*), то разрядность коэффициентов придется округлить до 53 бит и тогда ошибка увеличится со значения  $4.33e-18$  для диапазона  $r [-(\ln 2)/2, (\ln 2)/2]$  до значения  $3.4e-16$ , что уже больше значения 1ulp формата *double*. Причем в том простом примере не учитывается влияние округления между операциями FMA внутри полинома, и очевидно, что чем больше степень полинома, тем больше цепь операций и больше накапливаемая ошибка. Этого можно избежать, применяя внутри расчетов арифметику типа *double-double* (*triple-double*, *quad-double*, если необходимо), которая заключается в представлении числа в виде суммы двух чисел с непересекающимися полями мантисс. Этот способ позволяет не округлять расчетную разрядность коэффициентов до 53 разрядов, а представлять каждый коэффициент в виде двух чисел формата *double*. Таким способом можно увеличить виртуальную разрядность внутренних вычислений до значений 100 и более битов.

### Применение компенсированных алгоритмов и переход к самосинхронной логике

Альтернативный способ — использовать компенсированные алгоритмы вычисления полинома по схеме Горнера, которые заключаются в вычислении ошибки вещественной операции параллельно основному расчету полинома. Использование такого компенсированного алгоритма обеспечивает уменьшение ошибки расчета до такого уровня, как если бы разрядность повысилась в 2 раза. Сравнение времени выполнения стандартной

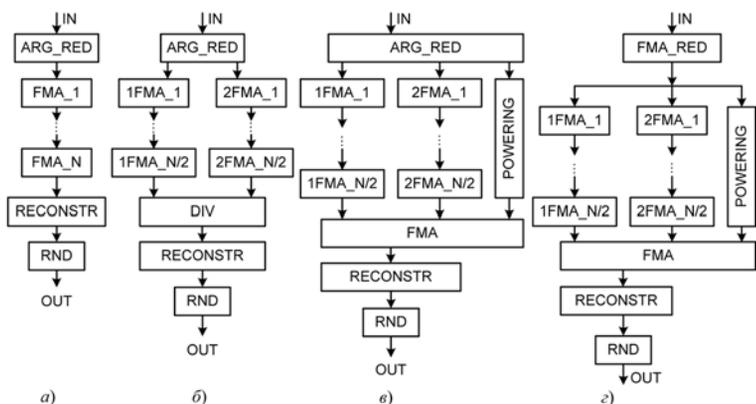


Рис. 8. Структурные схемы конвейерных алгоритмов вычисления элементарных функций

схемы Горнера, схемы Горнера с использованием арифметики *double-double* и компенсированной схемы Горнера показывает, что применение схемы *double-double* обеспечивает замедление по сравнению со стандартной схемой в 8 раз, в то время как применение компенсированной схемы — замедление в 3 раза [13, 14].

В этом случае вычисление значения полинома происходит на отдельных операциях умножения и сложения, а ошибка операций вычисляется с использованием блоков FMA:

$function [x, y] = FastTwoSum(a, b)$   
 $function [x, y] = TwoProd(a, b)$

$|a| \geq |b|: |a| < |b|: x = a * b$

$x = a + b \quad x = a + b \quad y = x - a * b = fma(x, a, b)$

$y = (a - x) + b \quad y = (b - x) + a$

Алгоритм компенсированной схемы Горнера будет выглядеть следующим образом:

$function r = CompHorner(P, x)$

$r_n = a_n; c_n = 0$

for  $i = n - 1 : -1 : 0$

$[p_i, \pi_i] = TwoProd(r_{i+1}, x)$

$t = c_{i+1} * x + \pi_i$

$[r_i, \sigma_i] = TwoSum(p_i, a_i)$

$c_i = t + \sigma_i$

$r = r_0 + c_0$

Применительно к архитектуре микропроцессора К64М с векторным сопроцессором вычисления по этой схеме могут быть частично распараллелены на блоке вещественной арифметики FPU и векторном сопроцессоре (рис. 9).

Описанные варианты схем вычисления полиномов предусматривают конвейерное исполнение и передачу данных от одной стадии к другой синхронно в соответствии с тактирующим сигналом или асинхронно по мере готовности результата на каждой стадии. Второй вариант исполнения придает всей схеме в целом характер потоковой вычислительной схемы. Один из вариантов реализации такой схемы основан на использовании блока FMA, выполненного как заказной блок в базе самосинхронной логики [15]. Такой вариант реализации не только дает большой выигрыш в энергоэффективности при сохранении быстродействия, сравнимого с синхронной реализацией, но повышает надежность схемы, включая функции самовосстановления после сбоя.

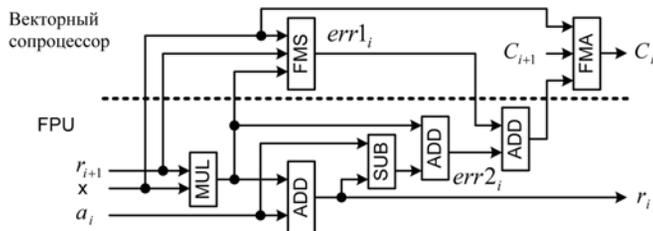


Рис. 9. Схема вычисления компенсированного алгоритма Горнера

Предложенная архитектура включает универсальное RISC-ядро микропроцессора и сопроцессоры, специализированные под выделенные функции, и позволяет разработать суперЭВМ эксафлопсного класса с использованием порядка  $10^6$  ядер, что на 2–3 порядка меньше необходимого числа универсальных ядер коммерческого применения без специализированных сопроцессоров. Алгоритмы, реализующие выделенные по результатам профилирования задачи функции, в целом решают поставленную задачу, но требуют дополнительного изучения задачи горения и оценки площади и производительности для окончательного принятия решения. Использование самосинхронной логики позволит решить ряд важнейших проблем, возникающих при разработке эксафлопсной суперЭВМ — снижение потребления и повышение надежности вычислительных узлов.

Работа выполнена в рамках проекта РФФИ № 13-07-12062 "Фундаментальные проблемы создания микропроцессоров и коммуникационных сред суперЭВМ эксафлопсного класса, ориентированных на предсказательное моделирование задачи горения".

#### Список литературы

1. Бетелин В. Б., Боксерман А. А., Костюков В. Е., Савельев В. А. Проблемы управления процессами повышения нефтеотдачи на основе моделирования на суперЭВМ // НефтегазоПромысловый ИНЖИНИРИНГ. 2010. № 3. С. 20–24.
2. Бетелин В. Б., Кушниренко А. Г., Смирнов Н. Н. Технологии супервычислений эксафлопсного класса и обеспечение конкурентоспособности энергетики России // НефтегазоПромысловый ИНЖИНИРИНГ. 2011. № 3. С. 10–13.
3. Рыбакин Б. П. Решение трехмерных задач газовой динамики на графических ускорителях // Вестник УГАТУ. 2012. Т. 16, № 6 (51).
4. Расширения набора команд для архитектуры Intel. URL: <http://software.intel.com/ru-ru/intel-isa-extensions>
5. Intel Xeon Phi Coprocessor Instruction Set Architecture Reference Manual. Intel Corporation, 2012.
6. Intel Xeon Phi product family performance brief. Rev.1.3, 10.04.2013. Intel Corporation, 2013.
7. Tuning the Intel MKL DFT functions performance on Intel® Xeon Phi™ coprocessors | Intel® Developer Zone. URL: <http://software.intel.com/en-us/articles/tuning-the-intel-mkl-dft-functions-performance-on-intel-xeon-phi-coprocessors>
8. Андрушин Д., Горбунов В., Эйсымонт Л. Перспективные особенности Tianhe-2 // Открытые системы. СУБД. 2013. № 8. С. 20–25.
9. Бобков С. Г., Аряшев С. И., Барских М. Е., Бычков К. С., Зубковский П. С. Микропроцессор гибридный: пат. 2359315 зарегистрирован в Государственном реестре изобретений Рос. Федерации 20.07.2009.
10. Зубковский П. С., Ивасюк Е. В., Аряшев С. И. Сопроцессор комплексных вычислений. Проблемы разработки перспективных микро- и наноэлектронных систем — 2010: сб. трудов. М.: Изд. ИППМ РАН, 2010. С. 356–359.
11. Muller J.-M. Elementary Functions: Algorithms and Implementation. Birkhauser, Boston, Basel, Berlin, 2005.
12. <http://lfpforge.ens-lyon.fr/www/crlibm/documentation.html>
13. Kornerup P., Lefèvre V., Louvet N. and Muller J.-M. On the computation of correctly-rounded sums // In Proc. of the 19th IEEE Symposium on Computer Arithmetic. Portland, OR, USA, June 2009. P. 155–160.
14. Langlois P., Louvet N. More Instruction Level Parallelism Explains the Actual Efficiency of Compensated Algorithms: Technical report. DALI research team, University of Perpignan, France 2007.
15. Плеханов Л. П. Основы самосинхронных электронных схем. М.: БИНОМ. Лаборатория знаний, 2013. 208 с.

С. Н. Шилов, аспирант, e-mail: shilov\_sn88@mail.ru,  
 С. Д. Кургалин, д-р физ.-мат. наук, зав. каф., e-mail: kurgalin@bk.ru,  
 А. А. Крыловецкий, канд. физ.-мат. наук, доц., e-mail: aakryl@cs.vsu.ru,  
 Воронежский государственный университет

## Анализ и реализация механизма репликации ресурсных записей в DNS-кластере

*Рассматривается актуальная проблема кластеризации информационных систем, в частности механизм репликации ресурсных записей в DNS-кластере. Рассмотрено два варианта реализации механизма репликации, используемых как часть системы распределения нагрузки. Представлен общий обзор системы распределения нагрузки, дано подробное описание и анализ реализованных репликационных механизмов.*

**Ключевые слова:** кластер, кластеризация, распределенная хеш-таблица, распределение нагрузки, балансировка нагрузки, репликация, система доменных имен

S. N. Shilov, S. D. Kurgalin, A. A. Krylovetsky

## Analysis and Implementation of Resource Records Replication Mechanism within DNS Cluster

*This article is devoted to the actual IT systems clustering problem, particularly to the resource records replication mechanism within DNS cluster. In scope of the article two versions of replication mechanism used as a part of the load distribution system are considered. The article provides general overview of the load distribution system, detailed description and analysis of implemented replication mechanisms.*

**Keywords:** cluster, clustering, distributed hash table, load distribution, load balancing, replication, domain name system

### Введение

В настоящее время кластеризация является одним из основных направлений развития информационных технологий. Применение кластеризации приводит к существенному повышению эффективности работы различных сервисов и систем.

Одной из наиболее широких областей применения кластеров являются сетевые сервисы. Кластеризация имеет ряд несомненных преимуществ по отношению к использованию отдельных высокопроизводительных машин. Прежде всего, это масштабируемость, высокая доступность сервиса (за счет использования множества узлов) и выгодное соотношение стоимости оборудования и суммарной производительности [3].

При реализации подобных систем актуальной проблемой является сохранение работоспособности системы при временных (по причине сетевого сбоя, технического обслуживания, обновления программного обеспечения, различного рода проблем в работе сервера и т. д.) или постоянных выходах узлов из состава кластера. Решением данной проблемы является репликация. При наличии соответствующих механизмов, системы, обеспечивающие функционирование кластера как единого целого,

позволяют организовать резервное копирование данных, которыми оперирует кластер. Применение подобных технологий существенно повышает надежность и стабильность кластеризованных систем.

Ранее в работе [1] нами были детально изложены принципы построения системы распределения нагрузки, основанной на технологии распределенной хеш-таблицы (*distributed hash table*, DHT), а также ее реализация. Представленная в работе [1] система предназначена для обеспечения функционирования кластера, составленного из рекурсивных кэширующих DNS-серверов. Помимо механизмов распределения нагрузки и обработки входа и выхода узлов из кластера, в системе реализована репликация ресурсных записей. Наличие этого процесса позволяет иметь несколько копий ресурсной записи (резервное копирование), что делает выход ответственного за эту запись узла абсолютно незаметным для пользователя.

Реализация механизма балансировки нагрузки подробно изложена в работе [1], в настоящей работе будут детально рассмотрены принципы, лежащие в основе механизма репликации ресурсных записей DNS (*domain name system* — система доменных имен).

## Особенности задачи

Проведем краткий обзор понятий и терминов, используемых в рамках данной задачи.

**Кластер** — это локально расположенная вычислительная система, состоящая из множества независимых компьютеров и связывающей их высокоскоростной сети.

Кластер состоит из трех основных компонентов: собственно обработчиков — компьютеров, образующих узлы кластера; сети, объединяющей эти узлы; программного обеспечения, заставляющего всю систему работать в стиле единого компьютера.

DNS — компьютерная распределенная система для получения информации о доменах. Чаще всего используется для получения IP-адреса по имени хоста (компьютера или устройства), получения информации о маршрутизации почты, обслуживающих узлах для протоколов в домене (SRV-запись). DNS необходима для работы Интернета, так как для соединения с узлом необходима информация о его IP-адресе.

DNS имеет следующие характеристики, имеющие большое значение при решении поставленной задачи:

- кэширование информации: узел может хранить некоторое количество данных не из своей зоны ответственности для уменьшения нагрузки на сеть;
- резервирование: за хранение и обслуживание своих узлов (зон) отвечают (обычно) несколько серверов, разделенных как физически, так и логически, что обеспечивает сохранность данных и продолжение работы даже в случае сбоя одного из узлов.

Также очень важным понятием системы DNS является запись DNS. Записи DNS, или ресурсные записи (англ. resource records, RR) — единицы хранения и передачи информации в DNS [4]. Именно этими записями оперируют DNS-серверы.

**Репликация** — резервное копирование данных для сохранения работоспособности или снижения влияния на функционирование распределенной системы.

Основные особенности задачи резервного копирования данных:

- резервное копирование должно выполняться на физически различные узлы кластера, в этом случае проблемы с оборудованием (дисковым, сетевым и т. д.) на конкретном узле не повлияют на общую работоспособность системы;
- резервное копирование должно осуществляться в соответствии с алгоритмами системы балансировки нагрузки, управляющей распределением входящих запросов на узлы кластера;
- резервное копирование не должно быть излишне избыточным для экономии ресурсов памяти и снижения нагрузки на локальную сеть, объединяющую узлы.

Также большое внимание должно уделяться сложности самого алгоритма репликации и связанном с ним накладным расходам, так как репликация не должна сильно влиять на производительность высоконагруженной системы, которой, в частности, является кластер DNS.

## Анализ проблемы

Одним из основных качественных показателей работы конкретного DNS-сервиса является скорость ответа. Абсолютно все DNS-провайдеры стремятся снизить время ожидания ответа клиентом. При этом, помимо особенностей сети, серверного оборудования и внутренней логики на уровне программного обеспечения, большое значение здесь имеет содержание и размер кэша ресурсных записей сервера (как упоминалось выше, подразумеваются рекурсивные кэширующие DNS-серверы). Наличие актуальной записи в кэше позволяет избежать обращения к вышестоящим и авторитативным серверам, которое может занимать достаточно большое время.

Объединение нескольких DNS-серверов в единый кластер позволяет получить распределенный кэш, при этом его общий размер является суммой размеров кэшей на каждом отдельном сервере. Распределенный кэш является одним из главных преимуществ, которое дает кластеризация в данном случае.

Однако распределенный кэш требует особого механизма обслуживания и поддержки для эффективного использования его содержимого. В особенности это касается масштабируемости кластера, ситуаций входа и выхода узлов. В основные задачи данной системы входит:

- балансировка нагрузки: система позволяет равномерно распределять входящие запросы между участниками кластера;
- обработка входа/выхода узлов: система позволяет узлам динамически входить и выходить из состава кластера без его остановки, при этом перестраивая свои внутренние структуры соответствующим образом.

Рассматриваемая система распределения нагрузки представляет собой DHT со специфичной реализацией согласованного хеширования (*consistent hashing*). В основе алгоритма распределения лежит 64-битная хеш-функция. Вся область значений используемой хеш-функции ( $2^{64}$ ) делится на  $2^{32}$  чанков, число чанков всегда является постоянным. В свою очередь, каждый чанк в равных частях делится на число "живых" узлов кластера. Таким образом, в каждом чанке конкретный узел кластера получает свою область ответственности. При изменении состава участников кластера область каждого чанка делится на новое число "живых" узлов.

Алгоритм выбора узла для обработки входящего запроса состоит в следующем: при поступлении запроса от запрашиваемого домена рассчитывается

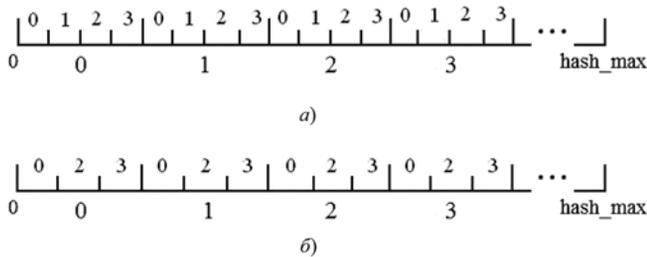


Рис. 1. Часть области значений хеш-функции до (а) и после (б) выхода узла 1 из состава кластера (под областью — номера чанков, над областью — обозначения областей ответственности узлов с соответствующими идентификаторами)

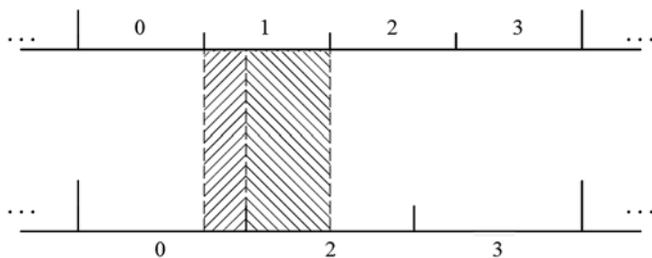


Рис. 2. Перераспределение области ответственности узла 1 в пределах одного чанка

хеш-значение, оно попадает в область ответственности какого-либо узла, который считается ответственным за данный домен, после чего запрос перенаправляется данному узлу на обработку.

Одно из главных требований к системе распределения нагрузки заключалось в том, что при выходе какого-либо узла из состава кластера его нагрузка (т. е. часть запросов, которые он обслуживал) должна равномерно перераспределиться между оставшимися участниками кластера. Если порядок следования узлов в каждом чанке будет одинаковым, то при выходе одного из узлов перераспределение нагрузки будет неравномерным. Рассмотрим в качестве примера узлы с идентификаторами 0, 1, 2 и 3. Представим графически часть области значений хеш-функции (рис. 1) до и после выхода узла 1 из состава кластера (рис. 1, а — часть области значе-

ний в линейном представлении до выхода узла 1; рис. 1, б — та же часть после выхода).

Как видно на рис. 1, а до выхода узла 1 из состава кластера пространство внутри каждого чанка делилось в равных частях между узлами 0, 1, 2 и 3. После выхода узла 1 произошло перераспределение пространств внутри чанков (рис. 1, б). Причем в каждом чанке область ответственности узла 1 (и, соответственно, часть обслуживаемых им запросов) перераспределилась только между узлами 0 и 2 (рис. 2).

Таким образом, перераспределение нагрузки произошло неравномерно и затронуло только узлы 0 и 2.

Для того чтобы при выходе узла из состава кластера его участки не переходили одним и тем же соседним узлам, порядок следования отрезков ответственности узлов должен быть неодинаковым (для равномерности перераспределения запросов при выходе узла из системы). Мы вводим 512 вариантов распределения узлов, повторяющихся циклично на всем пространстве [1]. Контейнером для хранения вариантов распределения является таблица распределения. Представим графически отображение таблицы распределения на область значений хеш-функции (рис. 3).

Таким образом, при выходе узла из состава кластера его области ответственности внутри различных чанков будут перераспределяться между всеми оставшимися участниками кластера. Число вариантов распределения должно быть более чем в 2 раза больше потенциального числа узлов, в противном случае число вариантов необходимо увеличить при запуске конкретного кластера. Неравномерное перераспределение области ответственности вышедшего узла внутри одного чанка (что видно на рис. 2) компенсируется большим количеством вариантов распределения и чанков, данный факт подтвержден экспериментально.

Очевидно, что содержание кэша конкретного узла зависит от областей его ответственности. Для поддержания актуальности кэшей узлы должны сохранять свое взаимное расположение в каждом из вариантов распределения при входе или выходе узлов, а также при добавлении новых участников кластера, что обеспечивается алгоритмом построения таблицы распределения. Подробное описание принципов и алгоритмов построения таблицы распределения представлено в работе [1].

Так как при выходе узла из состава кластера его области ответственности перераспределяются среди оставшихся активных узлов, то возникает ситуация, когда узлы начинают получать новые для себя домены, которые они не обслуживали ранее. Соответственно, эти домены отсутствуют в локальных кэ-

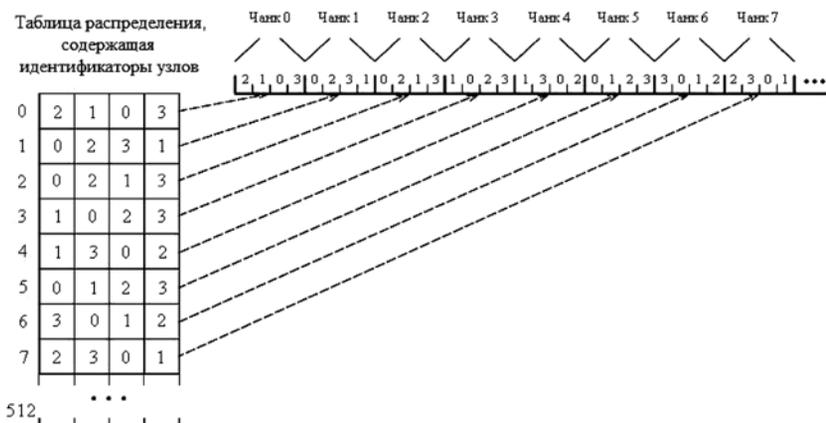


Рис. 3. Отображение таблицы распределения на область значений хеш-функции

шах оставшихся узлов, что заставляет их чаще обращаться к авторитативным серверам. Это приводит к временному снижению производительности кластера, ему требуется некоторое время на стабилизацию работы.

Чтобы избежать подобной ситуации, был реализован механизм репликации ресурсных записей, позволяющий хранить несколько копий записи на различных узлах кластера, причем выбор узлов для репликации соответствует алгоритму распределения запросов. Алгоритм репликации должен "предугадывать", какие узлы будут обслуживать данный домен при выходе ответственного узла, чтобы репликация была эффективной и не излишне избыточной.

### Реализация

В результате анализа проблемы проектирования и разработки алгоритмов было реализовано две версии механизма репликации. Рассмотрим подробно оба решения.

Первый вариант механизма репликации основывается на принципе пересечения областей ответственности узлов. Основная идея состоит во взаимном перекрытии отрезков ответственности, в соответствии с которой определяется не только узел, отвечающий за данный участок хеш-пространства, но и узел, чей отрезок имеет с ним зону перекрытия. При этом первому из них будет послан запрос на разрешение домена, а второму — специальный запрос от первичного узла после разрешения домена, содержащий соответствующие ресурсные записи, только получив запрос узел выполнит кэширование данных.

Размер области перекрытия является конфигурируемым параметром, который задается в процентном соотношении перекрытия. Значение перекрытия задается при старте системы, а также может быть переопределено уже во время ее функционирования, что существенно добавляет гибкости управлению системой.

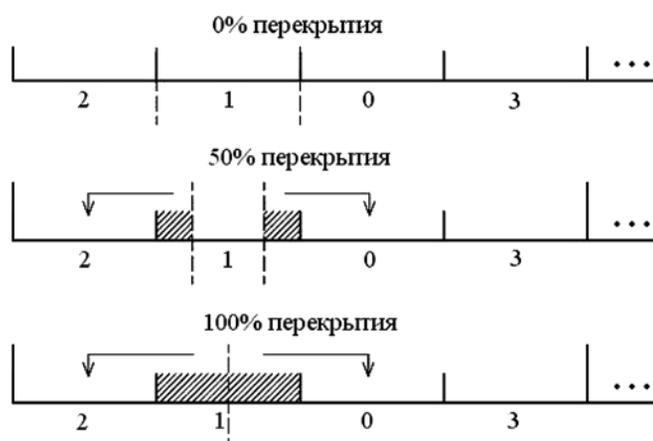


Рис. 4. Репликация записей на основе взаимного перекрытия областей ответственности узлов

На рис. 4 изображены три варианта перекрытия областей для узла с идентификатором 1 в пределах одного чанка с определенным распределением узлов: 0 % перекрытия (без репликации), 50 % и 100 % перекрытия. Здесь при попадании ключа в заштрихованные области запрос на разрешение домена будет послан узлу 1, а соответствующему перекрывающему узлу (с идентификатором 2 или 0) будет послан специальный запрос на кэширование данных. При попадании ключа в незаштрихованную область ответственности узла 1 будет послан только один запрос на разрешение домена, репликация в этом случае выполняться не будет. Как видно на рис. 4, при 100 % перекрытия репликация будет проводиться при поступлении каждого запроса. В изображенном чанке ключи из области ответственности узла 1 будут реплицироваться узлам с идентификаторами 2 и 0, но так как в других чанках будут иметь место различные варианты распределения узлов, то и реплицироваться ключи из областей узла 1 будут различным узлам системы. В этом смысле также присутствует равномерность распределения нагрузки на узлы кластера.

Представим пошагово рассмотренный алгоритм (перед началом работы алгоритма репликации мы уже имеем хеш-значение для домена, посчитанное системой распределения нагрузки).

*Шаг 1.* По хеш-значению определяется номер чанка и его идентификатор от 0 до 511 (по числу вариантов распределения).

*Шаг 2.* Определяется ответственный узел и положение ключа в данном чанке как остаток от деления хеш-значения на размер чанка.

*Шаг 3.* По положению ключа в чанке рассчитывается его положение уже относительно границ области ответственности узла как остаток от деления на размер области ответственности.

*Шаг 4.* Проводится проверка, попадает ли ключ, положение которого было найдено на предыдущем шаге, в заданные области перекрытия справа и слева.

*Шаг 5.* В случае попадания в область перекрытия соответствующий соседний узел выбирается для репликации ресурсных записей запрашиваемого домена, иначе репликация выполняться не будет.

Таким образом, при выходе узла из кластера перекрывающие его узлы уже будут частично содержать его кэш, что положительно скажется на производительности кластера.

Однако такое решение имеет существенный недостаток. В случае, когда процент перекрытия пространства меньше 100 %, невозможно заранее предугадать, какие домены будут реплицироваться, а какие нет. Это вносит неясность в поведение репликационного механизма, оно становится несколько непредсказуемым. Также при таком подходе мы не можем получить больше, чем одну копию данных. Несколько копий на различных узлах не поддерживаются в рамках данного алгоритма.

В результате усовершенствования описанного решения был реализован второй вариант механизма репликации, в котором удалось избавиться от недостатков первого.

Основная идея состоит в том, чтобы всегда проводить репликацию на четное число узлов, которые потенциально могут стать ответственными за данный домен. В самом простом случае репликация будет выполняться на два соседних узла. Стоит заметить, что в соответствии с алгоритмом распределения нагрузки, репликация должна проводиться только в пределах чанка. То есть, если ответственный узел находится в чанке на крайней позиции (первым или последним), то для репликации будут выбраны последующие (или предыдущие) узлы.

Распределенная система хранения данных *Dynamo* от *Amazon*, где также используется принцип согласованного хеширования (*consistent hashing*), имеет несколько похожий алгоритм репликации [2]. Однако рассматриваемый репликационный механизм реализован с учетом принципов, лежащих в основе системы распределения нагрузки, вследствие чего имеют место существенные отличия от алгоритма репликации системы *Dynamo*.

На рис. 5 изображены два случая для варианта репликации с двумя резервными копиями для четырех узлов кластера. Заштрихованная область —

область ответственности узла, в которую попадает запрашиваемый домен. Стрелками указаны узлы, которые будут выбраны для репликации записей запрашиваемого домена. В первом случае (рис. 5, а) домен попал в область ответственности узла 1, репликация будет проводиться на соседние для него узлы 2 и 0. Во втором случае (рис. 5, б) запрашиваемый домен попал в область узла 3, который является крайним в рассматриваемом чанке. Следовательно, репликация в данном случае будет проводиться на два предыдущих узла 0 и 1.

Отметим принципиальную важность четности числа узлов для репликации. В случае, когда узлами для репликации выбираются соседние узлы по отношению к ответственному, достаточно не просто заранее предугадать, какой именно из узлов будет обслуживать конкретный реплицированный домен после выхода ответственного узла.

Представим конкретную ситуацию на примере рассмотренного выше чанка. Пусть ответственным для доменов *example1.com* и *example2.com* является узел 1. Соответственно, *example1.com* и *example2.com* будут реплицированы на узлы 2 и 0. Однако при выходе узла 1 из состава кластера домены *example1.com* и *example2.com* уже будут обслуживаться различными узлами в соответствии с их областями ответственности (рис. 6).

Как видно на рис. 6, после выхода узла 1 домен *example1.com* будет обслуживаться узлом 0 (рис. 6, а), а домен *example2.com* — узлом 2 (рис. 6, б). Для того чтобы до выхода узла 1 определить конкретный узел, который будет обслуживать *example1.com* или *example2.com*, и выполнить репликацию только на него, необходимо вычислить будущую границу между областями ответственности узлов 2 и 0. Причем такое вычисление придется делать при каждом входящем запросе, что отрицательно скажется на производительности алгоритма балансировки нагрузки. Так как быстродействие является здесь важнейшим показателем, предлагается ввести некоторую избыточность репликации и выполнить ее на четное число узлов.

Однако здесь нельзя говорить об избыточности в полной мере, так как при одновременном выходе из кластера узлов 0 и 1 оба домена (*example1.com* и *example2.com*) будет обслуживать узел 2, который уже будет содержать эти домены в кэше благодаря репликации (несмотря на то что до выхода узлов 0 и 1 репликация домена *example1.com* на узел 2 могла показаться избыточной).

Также алгоритм позволяет задавать уровень покрытия репликации, т. е. число резервных копий: 2, 4, 6 и т. д. На рис. 7 представлен пример с четырьмя резервными копиями в случае кластера из семи узлов.

Таким образом, предложенный алгоритм позволяет задавать максимальный уровень репликации с полным покрытием, при котором каждый узел

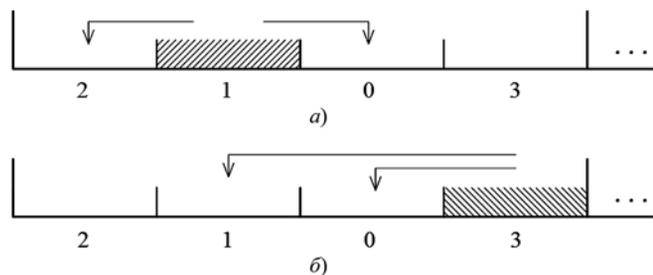


Рис. 5. Репликация записей с двумя резервными копиями для кластера из четырех узлов

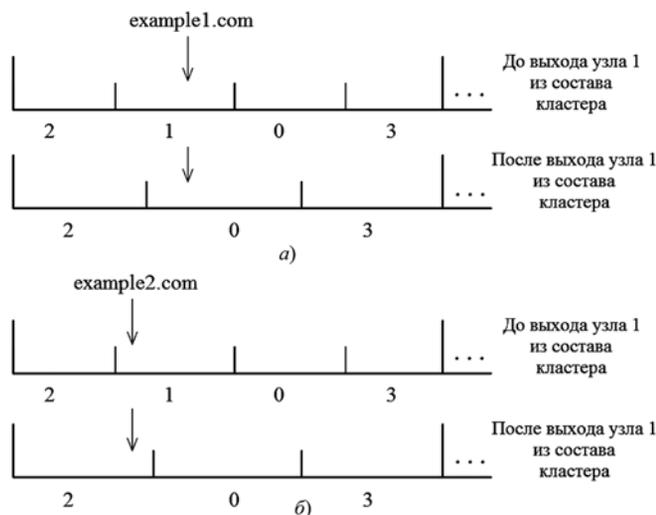


Рис. 6. Обслуживание доменов *example1.com* (а) и *example2.com* (б) до и после выхода узла 1 из состава кластера

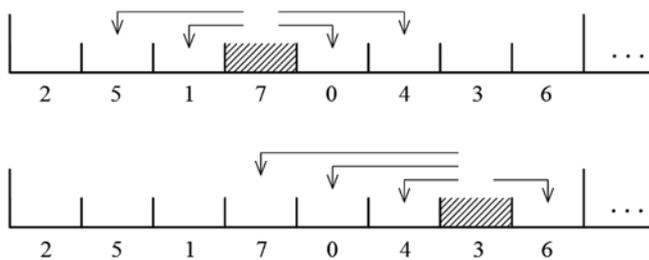


Рис. 7. Репликация записей с четырьмя резервными копиями для кластера из восьми узлов

будет содержать полную копию ресурсных записей других узлов кластера.

Рассмотрим алгоритм пошагово (как и в первом варианте перед началом работы алгоритма мы уже имеем хеш-значение для домена).

**Шаг 1.** По рассчитанному хеш-значению определяется номер чанка и его идентификатор от 0 до 511.

**Шаг 2.** Определяется ответственный узел на основании идентификатора чанка, хеш-значения и размера областей ответственности узлов.

**Шаг 3.** По таблице распределения выбирается указанное число соседних узлов внутри заданного чанка, стоящих последовательно справа и слева от ответственного узла, с учетом возможных описанных выше краевых условий.

**Шаг 4.** На выбранные узлы проводится репликация.

## Заключение

Представляется, что второй рассмотренный вариант с репликацией на четное число узлов выглядит гораздо более предпочтительным, чем первый, где репликация проводится на основе перекрытия участков ответственности. Второй вариант гораздо более предсказуем и управляем. Также он позволяет иметь множество копий ресурсных записей, вплоть до резервного копирования на все узлы кластера. К настоящему времени именно этот вариант нами был успешно реализован для ряда DNS-кластеров в Европе и США. Применение данного принципа репликации позволило существенно повысить стабильность работы кластера при плановом обслуживании, обновлении программного обеспечения, сбоях в работе серверов и т. д. Благодаря репликации выход какого-либо узла из состава кластера стал абсолютно незаметным для пользователя.

Как было указано выше, второй вариант реализации имеет много общего с механизмом репликации системы *Dynamo* от *Amazon* [2]. Однако существуют свои особенности, связанные с конкретной реализацией системы распределения нагрузки на узлы кластера. Провести полноценное сравнение этих систем на данный момент практически невозможно в связи с тем, что система *Dynamo* является закрытым решением, известны лишь общие прин-

ципы, описанные в работе [2]. Исходя из этой информации, можно заключить, что обе системы опираются на взаимное расположение отрезков ответственности узлов на области значений используемой хеш-функции при выборе узлов для репликации. Разница состоит в том, что *Dynamo* выбирает  $N - 1$  узел по часовой стрелке (область значений хеш-функции в этом случае для удобства представляется в виде кольца, где нулевое значение "соединяется" с максимальным) [2]. Таким образом, обе системы способны покрыть репликацией произвольное число узлов, используя при этом несколько разные алгоритмы, соответствующие сценариям обработки выхода узлов. Данные о производительности *Dynamo* в части репликации не опубликованы. Однако о рассмотренной реализации можно сказать, что при составлении списка узлов для резервного копирования алгоритм выполняет лишь несколько арифметических действий и смещений в массиве. Набор этих операций выполняется крайне быстро (менее 1 мкс).

Если говорить об аналогичных системах в области DNS, то на данный момент подобные кластерные решения являются абсолютно закрытыми разработками ведущих DNS-провайдеров, информация о которых отсутствует в открытых источниках.

Тестирование в лабораторных условиях и реальная эксплуатация второго варианта реализации репликационного механизма показали полное соответствие предъявляемым требованиям в области надежности и быстродействия. Но первый вариант также может давать преимущества в определенных ситуациях, например, при существенно ограниченных ресурсах памяти, вследствие чего размеры локальных кэшей могут быть сильно лимитированы, и репликация на четное число узлов может стать слишком "дорогой". В таком случае подобная "частичная" репликация поможет повысить стабильность работы кластера при небольших дополнительных затратах памяти на хранение некоторого процента реплицированных ресурсных записей.

## Список литературы

1. Шилов С. Н., Кургалин С. Д., Крыловецкий А. А. Реализация инфраструктуры распределенной хеш-таблицы в рамках кластерной системы DNS // Вестник ВГУ. Серия: Системный анализ и информационные технологии. 2012. № 2. С. 74–79.
2. DeCandia G., Hastorun D., Jampani M., Kakulapati G., Lakshman A., Pilchin A., Sivasubramanian S., Vosshall P., Vogels W. Dynamo: Amazon's Highly Available Key-value Store // In Proc. of the 21<sup>st</sup> ACM Symposium on Operating Systems Principles (Skamania Lodge Stevenson, WA, USA, October 14–17, 2007). New York: ACM Press, 2007. P. 205–218.
3. Fox A., Gribble S. D., Chawathe Y., Brewer E. A., Gauthier P. Cluster-based scalable network services // In Proc. of the Sixteenth ACM Symposium on Operating Systems Principles (Saint Malo, France, October 05–08, 1997). New York: ACM Press, 1997. P. 78–91.
4. Mockapetris P. Domain Names — Concepts and Facilities // RFC 1034, The Internet Society, November 1987.

# МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ MODELING AND OPTIMIZATION

УДК 519.17

А. О. Аристов, канд. техн. наук, доц. каф. "Системы автоматизированного проектирования"  
Московского государственного горного университета, e-mail: batan-87@mail.ru

## Особенности моделирования потоковых систем на основе квазиклеточных сетей с использованием структурной методологии проектирования

*Рассмотрены вопросы проектирования и разработки инструментариев моделирования на основе квазиклеточных сетей, представляющих собой новый тип динамических дискретных структур, не имеющих сигнатуру и позволяющих моделировать и проектировать поточные системы. На основе стандарта IDEF0 построена модель, являющаяся организационным обеспечением процесса проектирования инструментариев моделирования квазиклеточными сетями при различных предметных интерпретациях. Внедрение квазиклеточных сетей рассмотрено на примере моделирования потоков посетителей объектов массового пребывания людей.*

**Ключевые слова:** квазиклеточные сети, дискретные структуры, проектирование, моделирование, сети, потоки

А. О. Aristov

## The Features of Quasi Cellular Nets in Simulations of Flow-Based Systems and Their Design with Structure Approach

*There is problems of design and development tools based on quasi cellular nets. Quasi cellular nets is new type of dynamical discrete structures without signature. It may be used for design and simulation flow systems. There was develop model based on IDEF0. There is organization models of quasi cellular network based processes modeling for different subjects. There is also samples of quasi cellular nets implementations in models flows of visitors on public objects especially stadium.*

**Keywords:** quasi cellular nets, discrete structures, design, simulation, nets, flows

### Квазиклеточные сети как перспективный подход к моделированию поточных систем

В работах автора [1–4] рассматривались квазиклеточные сети как фундаментальные дискретные структуры, не имеющие сигнатуры [1, 3]. Особенности функционирования квазиклеточных сетей тесно связаны с понятием циркуляции [1, 3], т. е. передачи состояния между соседними клетками. Переход бинарного состояния между соседними клетками приведен на рис. 1. Введена широкая клас-



Рис. 1. Переход бинарного состояния между клетками

сификация квазиклеточных сетей [2, 5]. Основным назначением квазиклеточных сетей является моделирование систем, рассмотрение которых сводится к моделированию потоков в сетях [2, 4, 5]. Речь идет о распространении каких-либо микрообъектов или частиц веществ в ограниченном пространстве.

Несмотря на то что квазиклеточные сети являются новым типом дискретных структур, они имеют некоторые сходства с широко известными графовыми моделями, сетями Петри, моделями потоков в сетях, клеточными автоматами и др. Следует отметить, что существуют различные методы синтеза квазиклеточных сетей [4]. В случае синтеза квазиклеточных сетей на основе базового графа графовая макромодель фактически приобретает свойства динамической дискретной структуры (рис. 2). Синтез методом битого клеточного автомата предполагает использование клеток из подмножества элементов

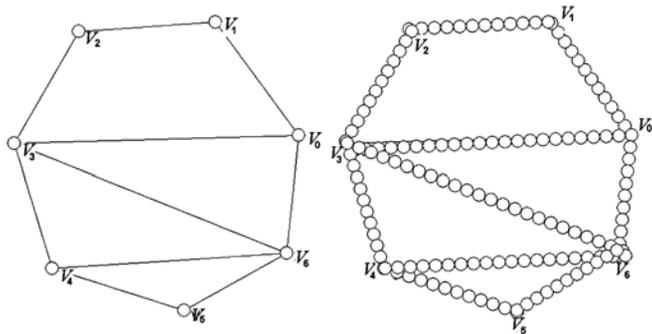


Рис. 2. Базовый граф и построенная на его основе квазиклеточная сеть

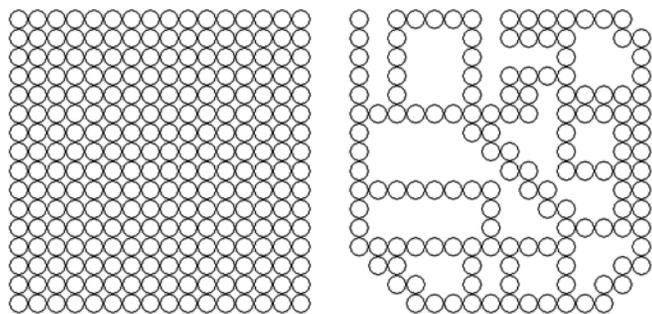


Рис. 3. Клеточный автомат и квазиклеточная сеть, синтезированная методом битого клеточного автомата

клеточного автомата (рис. 3). Также следует отметить возможности синтеза квазиклеточных сетей на базе поведения микрообъекта, в каждой координатах которого формируется клетка [4].

Таким образом, сравнивая квазиклеточные сети с другими подходами к моделированию систем, предполагающих распространение потоков, в частности, по сравнению с использованием теоретико-графовых моделей, описанных в работах Форда и Фалкерсона [6], а также микромоделей, описанных в ряде работ по имитационному компьютерному моделированию [7, 8 и др.], нетрудно видеть, что квазиклеточные сети позволяют реализовать моделирование потока на микро- и макроуровнях в рамках единой дискретной структуры. Подобные возможности можно наблюдать также в клеточных автоматах, но в отличие от них квазиклеточные сети не требуют периодического разделения пространства (сравнимого с растеризацией), а предполагают выделение дискретных областей на путях распространения потоков.

### Структурное проектирование программных средств моделирования систем на основе квазиклеточных сетей

Следует однако обратить внимание на то, что фундаментальные дискретные структуры в чистом виде не применяют, а их приложение сводится к нахожде-

нию предметных интерпретаций указанных структур, их элементов и протекающих в них явлений [9]. Поэтому при разработке моделей на основе квазиклеточных сетей следует определить, как будут интерпретироваться в различных предметных областях такие понятия, как клетка, состояние, передача состояния, циркуляция, источник, приемник, клеточный автомат, турникет, анастомоз и др. Интерпретация указанных понятий в различных предметных областях позволит применить квазиклеточную сеть для моделирования и представления моделируемых систем [3, 4].

Рассмотрим общую модель применения теории квазиклеточных сетей для решения задачи моделирования системы в различных предметных областях. Модель рассмотрим в соответствии со стандартом IDEF0 (ГОСТ Р 50.1.028—2001).

Предполагается, что исходными данными для разработки моделей на основе квазиклеточных сетей является предметное описание моделируемой системы, включающее описание элементов системы, их поведения, структуры системы и особенностей ее функционирования (рис. 4). Следует отметить, что речь идет о системах, функционирование которых предполагает наличие потоков, распространяемых в некотором пространстве, определяемом линейными, угловыми, временными и другими координатами.

Результатом моделирования системы на основе квазиклеточных сетей являются экспериментальные данные, полученные в процессе функционирования квазиклеточной сети и используемые для дальнейшей обработки, анализа, поддержки принятия решения и т. д. [2]. Также следует отметить, что в настоящее время ряд прикладных проектов в сфере информационных технологий сводится не к получению конкретных данных, а к разработке инструментариев, ориентированных на использование специалистами в предметных областях. В такой ситуации решается не предметная задача, связанная



Рис. 4. Контекстная диаграмма процесса моделирования на основе квазиклеточных сетей



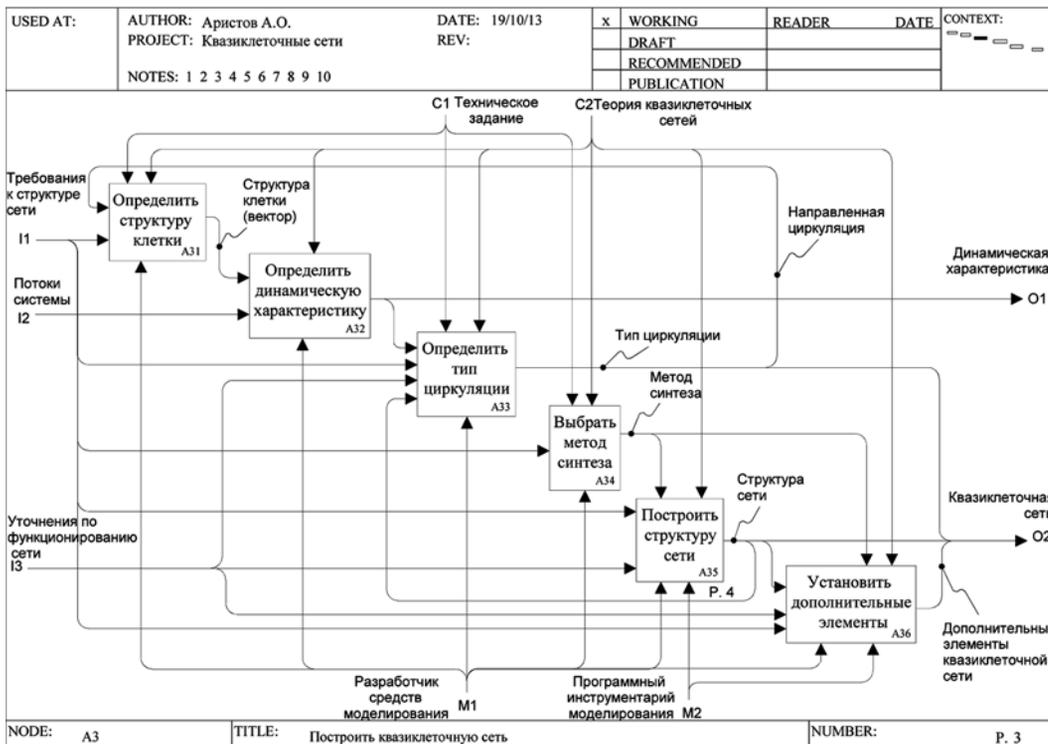


Рис. 6. Проектирование квазиклеточной сети

где  $B_i$  — неизменные (базовые) параметры клетки (от англ. *Basic*);  $C_i$  — параметры клетки, изменяющиеся при прохождении фишек через клетку (от англ. *Changeable*);  $S_i$  — параметры фишки как объекта, находящегося в клетке, т. е. переменные состояния (фазовые переменные) клетки (от англ. *State*). На первом этапе проектирования квазиклеточных сетей определяется набор параметров (блок A31), образующих клетку в соответствии с (1). В структуру клетки входят фазовые переменные, определяющие состояние. Динамический аспект квазиклеточных сетей предполагает изменение состояния с течением времени (или при каждом такте изменения динамической характеристики). В соответствии с динамической характеристикой осуществляется циркуляция различного типа — случайная, направленная, микроциркуляция и др. Те или иные типы циркуляции позволяют моделировать различные варианты распространения потоков в квазиклеточных сетях.

Следующий этап проектирования квазиклеточных сетей — синтез структуры, предполагающей расположение клеток в пространстве и фактически влияющей на распространение потоков и циркуляцию. Выше отмечалось несколько методов синтеза, среди которых метод базового графа, метод генерирующей фишки и метод битого клеточного автомата.

Метод базового графа позволяет преобразовать статическую модель системы, заданную в виде графа в динамическую модель (квазиклеточную сеть). Метод битого клеточного автомата позволяет до-

бавлять элементы клеточных автоматов и их подмножеств в состав квазиклеточных сетей. Метод генерирующей фишки основан на поведении микрообъекта. Такой метод синтеза целесообразно использовать, когда известна математическая модель движения множества однотипных объектов. Таким образом, при различных предметных интерпретациях используются те или иные методы синтеза квазиклеточных сетей. Завершающим этапом синтеза структуры квазиклеточной сети является установка дополнительных компонентов, таких как генераторы (истоки), стоки и турникеты.

Отдельно следует отметить, что рассмотренные методы синтеза квазиклеточных сетей могут использоваться совместно. Указанные особенности отражены на диаграмме декомпозиции процесса синтеза (рис. 7).

Следует также отметить, что процесс синтеза квазиклеточных сетей предполагает возможность многократного повторения отдельных этапов. Во многом это связано с необходимостью уточнения и изменения структуры сети. Также следует отметить, что ряд задач моделирования и поиска проектного решения сводится к варьированию параметров модели, что также предполагает возврат к предыдущим этапам [10]. Кроме того, особенностью квазиклеточной сети является тот факт, что значительная часть параметров квазиклеточной сети вычисляется в ходе программной реализации. Это связано с нецелесообразностью проведения расчетов циркуляции вручную. Исходя из этого, на этапе синтеза квазиклеточной сети предполагается использование про-



тами клеточных автоматов. Второй вариант — построение методом битого клеточного автомата. Метод битого клеточного автомата в любой ситуации позволяет моделировать какие-либо препятствия, например установление дополнительных заграждений на стадионе. Заключительным этапом проектирования структуры квазиклеточной сети является установка турникетов. Истоки, стоки и измерительный участок квазиклеточной сети определяются исходя из целей моделирования [11].

Пример оцифровки проходной зоны стадиона представлен на рис. 10 (см. вторую сторону обложки). Следует отметить, что квазиклеточная сеть формируется на основе нескольких областей, представляющих собой вкрапление элементов клеточного автомата [11]. Предметно области клеточного автомата следует рассматривать как зоны с возможностью свободного перемещения посетителей, однако в случае движения по определенным траекториям, например в условиях движения вдоль заграждений, базовый граф примет вид, представленный на рис. 11. Кроме того, следует представить конфигурацию базового графа для случая общей очереди и ее обслуживания на нескольких пропускных пунктах (рис. 12). Также примером подобной ситуации можно считать обслуживание общей очереди людей несколькими кассами.

Таким образом, структура квазиклеточной сети задает ограничения для координат движущихся объектов, в частности людей на объектах массового прибывания.

На следующем этапе необходимо определить структуру клетки квазиклеточной сети [4]. Каждая клетка задается в виде (1).

В задаче моделирования поведения посетителей учреждений массового пребывания людей в качестве базовых параметров выступают координаты единицы пространства (клетки) на плоскости:

$$B_i = (x_i, y_i, z_i), \quad (2)$$

при этом для многоуровневого объекта рассматривается трехмерное распределение единиц пространства, где  $z_i$  — уровень (этаж).

В качестве вектора состояния следует рассматривать набор параметров каждого посетителя, находящегося в указанной области пространства (клетке). Однако следует отметить, что ряд параметров связан со значениями элементов вектора  $B_i$ , в частности координаты, в которых находится посетитель. Учитывая, что движение посетителя может быть неравномерным, а также не всегда проходит через центр клетки, моделирующей область пространства, реальные координаты посетителя  $(x_i^{(p)}, y_i^{(p)})$  отклоняются на случайную величину:

$$\begin{cases} x_i^{(p)} = x_i + \text{random} \cdot 2R - R \\ y_i^{(p)} = y_i + \text{random} \cdot 2R - R, \end{cases} \quad (3)$$

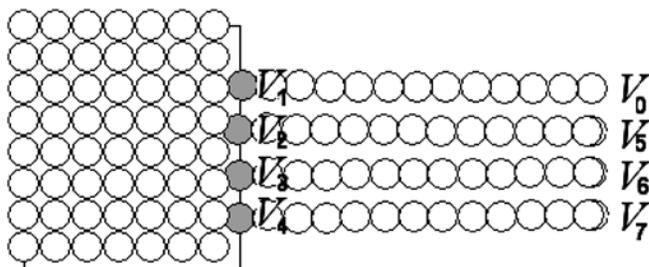
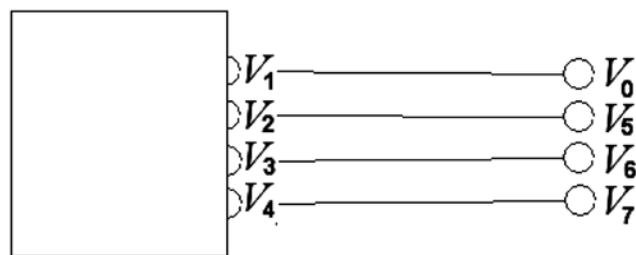


Рис. 11. Базовый граф и квазиклеточная сеть для движения с заграждениями

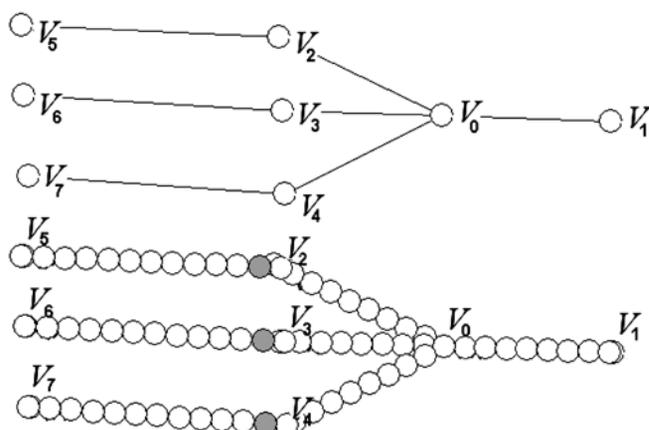


Рис. 12. Базовый граф и квазиклеточная сеть для обслуживания общей очереди

где  $\text{random}$  — величина, принимающая случайное значение, равномерно распределенное в диапазоне  $[0,1]$ ;  $R$  — область пространства. Следует отметить, что

$$(x_i^{(p)}, y_i^{(p)}) \in S_i, \quad (4)$$

Координаты посетителя  $(x_i^{(p)}, y_i^{(p)})$  вычисляются каждый момент времени исходя из того, в какой клетке находится посетитель. Несмотря на отсутствие необходимости представления координат в составе вектора параметров  $S_i$ , указанный вектор содержит ряд других параметров. В простейшем случае в качестве таких параметров следует рассматривать наличие посетителя в заданной области пространства. Факт наличия посетителя кодируется бинарным значением:

$$S_i \in (0, 1). \quad (5)$$

В зависимости от размера области  $R$  наличие фишки в клетке интерпретируется как наличие од-

ного или нескольких человек в области пространства, тогда состояние приобретает вид вектора бинарных компонент.

Для класса задач, в которых воздействием посетителей на сам объект можно пренебречь, вектор переменных параметров остается пустым, однако в случае, когда необходима статистика, связанная с прохождением посетителей через измерительный участок, в качестве вектора параметров выступают величины, являющиеся счетчиками, т. е.

$$C_i(t + \theta) = C_i(t) + 1 \text{ при } \begin{cases} S_i(t) = 0 \\ S_i(t + \theta) = 1. \end{cases} \quad (6)$$

Значение счетчика (7) используется при вычислении ряда значений путем моделирования потоков посетителей объектов массового нахождения людей.

Таким образом, определена структура клетки:

$$Q_i = (x_i, y_i, z_i, C_i, S_i), \quad (7)$$

где  $x_i, y_i$  — координаты посетителя;  $z_i$  — уровень (этаж);  $C_i$  — счетчик людей, проходящих через область пространства;  $S_i$  — наличие посетителей в области пространства.

Моделирование посетителей на объектах массового пребывания людей осуществляется с течением модельного времени. Фактически с течением времени меняются значения счетчика и состояние, задаваемые численно, тогда:

$$Q_i = (x_i, y_i, z_i, C_i(t), S_i(t)). \quad (8)$$

Следующим этапом проектирования моделей на основе квазиклеточных сетей является выбор типа циркуляции. Для объектов массового пребывания людей характерны различные типы циркуляции, в первую очередь зависящие от специфических особенностей решаемой задачи. Поскольку в любом случае предполагается движение потоков посетителей в определенном направлении, то целесообразно использовать направленную циркуляцию, а также микроциркуляцию. Причем микроциркуляция используется в тех ситуациях, когда маршруты перемещений заранее известны [5]. В случаях, когда моделируется прохождение посетителей и оценивается время прохождения, используется направленная циркуляция, поскольку для таких объектов характерно передвижение посетителей к некоторым целевым точкам. Пусть посетители объекта массового пребывания перемещаются в направлении некоторой точки с координатами  $(x_T, y_T)$ . Предметно в виде такой точки представляются турникеты, выходы из залов, места в зале и т. п. Для случая моделирования проходного пункта стадиона в качестве такой точки целесообразно рассматривать входы на трибуну. Тогда при нахождении посетителя объекта в точке  $Q_c = (x_c, y_c, z_c, C_c(t), S_c(t))$  при наличии клеток

$$Q_{nj} = (x_{nj}, y_{nj}, z_{nj}, C_{nj}(t), S_{nj}(t)), \quad (9)$$

где  $j = 1, 2, \dots$ , для каждой из которых выполняются условия

$$\begin{cases} (x_{nj} - x_c)^2 + (y_{nj} - y_c)^2 \leq 4R^2 \\ S_{nj}(t) = 0, \end{cases} \quad (10)$$

будет осуществляться переход  $Q_c \rightarrow Q_{nj}$ , при котором

$$\min_j \sqrt{(x_{nj} - x_T)^2 + (y_{nj} - y_T)^2}. \quad (11)$$

Таким образом, решены вопросы формирования модели пространства объекта массового пребывания людей для моделирования поведения посетителей на основе квазиклеточных сетей. Обосновано применение направленной циркуляции либо микроциркуляции [5].

После построения квазиклеточной сети необходимо определить участок, на котором будут выполняться измерения. Для оценки числа посетителей, проходящих через проходную зону стадиона, установим измерительный участок на турникетах. Кроме того, следует отметить, что в структуру каждой клетки включены счетчики  $C_i(t)$ , показывающие число посетителей, прошедших через указанную клетку за время  $t$ . Тогда на измерительном участке  $Q = \{Q_1, Q_2, \dots, Q_n\}$

$$C(t) = \sum_{i=1}^n C_i(t). \quad (12)$$

Таким образом, оценка числа посетителей через турникеты объекта массового пребывания людей является аддитивным макропараметром измерительного участка квазиклеточной сети [5, 8].

Еще ряд величин был рассмотрен ранее и их нахождение выполняется на основе формул, рассмотренных в работах [5, 8]. Указанные величины фактически не зависят от интерпретации элементов квазиклеточных сетей.

\*\*\*

Таким образом, в рамках данной работы представлена функциональная модель, описывающая процесс проектирования, разработки и внедрения моделей систем на базе квазиклеточных сетей. Как отмечалось выше, квазиклеточные сети представляют собой фундаментальные дискретные структуры, позволяющие моделировать на микро- и макроуровнях системы, предполагающие рассмотрение потоков объектов при различных предметных интерпретациях. В качестве примера применения квазиклеточных сетей приведены объекты массового пребывания людей. Функциональная модель позволяет организовать процесс разработки и внедрения средств моделирования для различных предметных интерпретаций квазиклеточных сетей.

### Список литературы

1. **Аристов А. О.** Квазиклеточные сети. Синтез и циркуляция // Горный информационно-аналитический бюллетень. 2013. № 2. С. 125—131.
2. **Аристов А. О.** Теория квазиклеточных сетей и ее приложения // II Международная научно-практическая конференция "Научно-техническое творчество молодежи — путь к обществу, основанному на знаниях": сб. науч. докладов. М.: Изд-во МГСУ, 2012. С. 230—234.
3. **Аристов А. О.** Квазиклеточные сети. Теоретическая база и программный инструментарий моделирования // Хроники объединенного фонда электронных ресурсов "Наука и Образование". 2012. № 11 (42). С. 25.
4. **Аристов А. О.** Методы синтеза квазиклеточных сетей // Научный вестник МГГУ. 2013. № 9 (42). С. 16—21.
5. **Аристов А. О.** Циркуляция в квазиклеточных сетях и их классификация // Горный информационно-аналитический бюллетень. 2013. № 9. С. 188—194.
6. **Форд Л. Р., Фалкерсон Д. Р.** Потоки в сетях: пер. с англ. М.: Мир, 1966. 277 с.
7. **Введение** в моделирование пешеходных потоков // Хабр. [электронный ресурс]. URL: <http://habrahabr.ru/post/158975/> дата обращения: 07.12.13.
8. **Аристов А. О.** Потоки в квазиклеточных сетях // Устойчивое инновационное развитие: проектирование и управление. Электрон. журн. 2013. № 3 (20). С. 36—41. URL: <http://www.gypravlenie.ru>.
9. **Горбатов В. А.** Фундаментальные основы дискретной математики. М.: Физматлит, 1999. 544 с.
10. **Аристов А. О., Моргачев К. В., Рябов Л. П., Суворов А. В., Федоров А. М.** Компьютерные системы поддержки принятия решений: учеб. пособие. М.: Изд-во МГГУ, 2012. 172 с.
11. **Аристов А. О.** Об элементах квазиклеточных сетей // Горный информационно-аналитический бюллетень. 2013. № 11. С. 322—332.

УДК 681.518:339.13

**Э. М. Димов**, д-р техн. наук, проф.,  
**О. Н. Маслов**, д-р техн. наук, проф., зав. каф., e-mail: [maslov@psati.ru](mailto:maslov@psati.ru),  
**Ю. В. Трошин**, канд. техн. наук, доц.,  
Поволжский государственный университет телекоммуникаций и информатики

## Снижение неопределенности выбора управленческих решений с помощью метода статистического имитационного моделирования

*Рассматривается проблема снижения неопределенности исходных данных при статистическом имитационном моделировании (СИМ) объектов с малой прецедентной базой. Показаны способы ее решения, направленные на повышение эффективности управленческих решений, принимаемых с помощью СИМ.*

**Ключевые слова:** метод статистического имитационного моделирования, неопределенность исходных данных, эффективность управленческих решений

A. M. Dimov, O. N. Maslov, Ju. V. Troshin

## Reducing Uncertainty in a Choice of Management Decisions Using Statistical Simulation

*The article is about problem of initial data uncertainty reducing in statistical simulation of objects with small case-base. There are shown solutions of that problem aimed at improving efficiency of management decisions made with the help of statistical simulation.*

**Keywords:** statistical simulation, initial data uncertainty, efficiency of management decisions

### Введение

Развитие теории систем управления (СУ) иерархическими и многокритериальными сложными системами (СС) организационно-технического типа (социально-экономическими, экологическими, военными и т. п.), неотъемлемыми компонентами которых являются лица, принимающие решения (ЛПР), до настоящего времени остается актуальной проблемой, имеющей важное практическое значение [1—5]. Анализ и синтез таких СУ встречает ряд

принципиальных трудностей: сложно, например, оптимизировать режим работы СС, компоненты которой (подсистемы и элементы) имеют возможность самостоятельно максимизировать свои функционалы. Поэтому, во-первых, в понятие оптимальности СУ, предназначенных для управления СС, "необходимо вложить конкретный разумный смысл" [6]. Во-вторых, следует рассмотреть перспективы использования при решении задач управления ими новых информационных технологий — в частности,

компьютерного метода статистического имитационного моделирования (СИМ) [2]. В-третьих, чтобы избежать путаницы в терминах, целесообразно сформулировать четыре следующих исходных тезиса.

1. Для достижения научно-исследовательских и коммерческих целей в настоящее время создаются "имитационные системы" [3, 4], каждая из которых представляет собой организационно-техническую (человеко-машинную, диалоговую) СС и состоит из математических (алгоритмических) моделей и системы процедур, позволяющей объединить эти модели с ЛПР, а также специального математического обеспечения, включающего систему алгоритмов для решения конкретных задач, в том числе оптимизационных.

2. Каждая такая имитационная (информационная) СС представляет собой наукоемкий проект, разработка и реализация которого требуют значительных усилий ЛПР разного профиля, базируются на результатах исследований фундаментального характера [6].

3. Проблема обеспечения адекватности СИМ связана с учетом как детерминированных характеристик, так и достаточно большого числа (до 100 и более) стохастических факторов (параметров СИМ-модели), влияющих на данный объект или процесс, проблема точности СИМ обусловлена тем, что полученные экспериментальные (числовые) результаты носят случайный характер [2, 3].

4. Возможности метода СИМ позволяют строить, наравне со статическими, динамические модели объектов (СС, СУ), эффективность функционирования которых зависит от взаимодействия во времени указанных стохастических факторов (случайных параметров СИМ-модели) [7, 8].

Анализ предметной области СИМ показывает, что случайность параметров является характерной особенностью моделей большинства реальных СС. Несмотря на то, что эти модели имеют градацию влияющих факторов: от наиболее важных (безусловно включаемых в состав модели) к неопределенным и незначительным (которыми ЛПР считают возможным пренебречь), воспользоваться данной градацией не всегда удастся, поскольку это сказывается на качестве (точности и адекватности) результатов СИМ. Наличие неопределенности параметров моделей СС и СУ, по нашему мнению, следует считать объективным неустранимым и неотъемлемым фактором, существенно влияющим на эффективность управления СС.

Цель статьи — анализ перспектив и возможностей СИМ по уменьшению влияния неопределенности параметров социально-экономических систем (далее СЭС) как одного из наиболее важных для практики вариантов реализации СС организационно-технического типа, СУ и их моделей (математических, компьютерных, имитационных) на качество управленческих решений. Принимая ответственность за отличия данной версии СИМ от

других известных аналогов [2, 3, 7—9], условимся именовать ее методом Димова—Маслова (МДМ). Главными особенностями МДМ являются:

- ориентация на слабоструктурируемые и плохо-формализуемые задачи, связанные с исследованием и управлением СЭС, с учетом трудности применения к ним методов и средств классической теории управления техническими СС [6];
- системный подход — ввиду многообразия свойств и рабочих характеристик СЭС, а также зачастую нестандартных и неоднозначных требований к их СИМ-моделям;
- использование знаний разного типа (верифицированных и аксиологических) при определении исходных условий для проведения СИМ.

### Принципы моделирования и требования к СИМ-модели

Введение СИМ-модели в состав СУ [4, 5] иллюстрирует рис. 1, где объектом управления является СЭС с входными параметрами  $X$  и выходными параметрами  $Y$ , реализующая бизнес-процесс  $Y = \Psi(X; U)$  в условиях возмущающих воздействий внешней (рыночной) среды  $E$ ; субъектом управления — ЛПР с управляющим воздействием  $U = F(Y; Y_M; E)$ ; при этом на вход СИМ-модели воздействуют входные параметры СЭС  $X$ , а на ее выходе фигурируют выходные параметры  $Y_M = \Psi(X; U; E)$ , которые поступают к ЛПР совместно с выходными параметрами СЭС как  $Y + Y_M$ . В работах [1, 2, 9] в общем виде, а в [10] применительно к условиям и целям функционирования конкретного объекта, представлена также двухконтурная схема СУ, где с помощью внешнего контура ЛПР может управлять СЭС помимо СИМ-модели, а внутренний контур включает СИМ-модель и предназначен для выбора наилучших в данный момент времени управленческих решений с учетом результатов, полученных с помощью МДМ (второй контур, см. более подробно на схеме рис. 3).

Считается, что способность ЛПР представлять взаимодействие частей СЭС менее надежна, чем знания о каждой из них. Поэтому, построив СИМ-

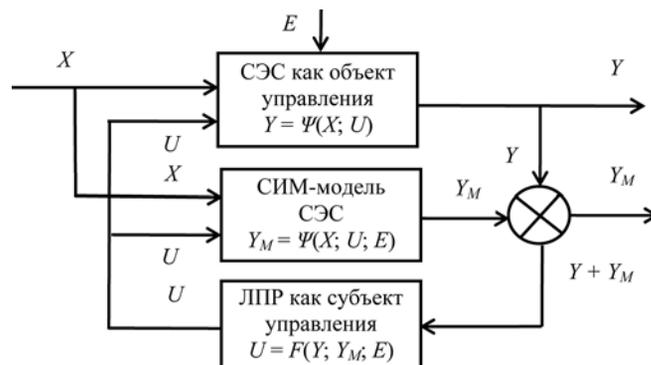


Рис. 1. Структура СУ с включением СИМ-модели СЭС в контур управления

модель и наблюдая на ней взаимодействие разных факторов, можно промоделировать и исследовать с помощью МДМ и возможные варианты состава, и структуры СЭС, и режимы ее работы с тем, чтобы выбирать наиболее подходящие из них в режиме реального времени. На СИМ-модель также возлагается задача выявления внешних факторов и возмущений, к которым чувствительна реальная СЭС; при этом модель может быть эффективной и в том случае, когда она правильно отражает лишь то, что мы посчитаем сущностью исследуемой СЭС после ее вербального моделирования. Запись функции решения [2, 7–9] предпочтительна в виде цепочки простых уравнений с учетом необходимых условий и ограничений. Отметим, что задача повышения точности вычислений не ставится, если СИМ-модель нечувствительна (робастна) к принятым допущениям (ошибки и искажения в рамках МДМ можно вносить преднамеренно, чтобы проверить это). Усложнение вычислительных методов делает формулировку уравнений менее понятной, тогда как преимуществами СИМ-модели, напротив, являются ее простота и наглядность [2–9, 11].

В то же время СИМ-модель призвана функционировать в максимально широких границах изменения переменных и параметров: во-первых, потому что в будущем может понадобиться расширение пределов изменения условий работы реальной СЭС; во-вторых, поскольку нельзя предсказать заранее, какие значения примут по ходу СИМ различные переменные; в-третьих, ввиду необходимости выяснить, как будет работать и окажется ли полезной СИМ-модель за пределами границ, которые имеются у рассматриваемой СЭС, поскольку разработка новых СЭС обычно предполагает их действие вне рамок прежней практики [2, 7, 8].

Если параметры и правила работы СЭС точно не определены и нет возможности их оценить, но они являются управляемыми, с помощью СИМ-моделей можно также исследовать воображаемые ситуации и сценарии развития событий несмотря на то, что речь будет идти о правдоподобии, а не о точности и адекватности (в строгом смысле слова) результатов СИМ по МДМ, предназначенных для управления СЭС.

### **Особенности проведения первоначальных этапов СИМ**

Методика СИМ для СЭС, подверженных воздействию  $E$  на рис. 1 (управляющему, исследователю, негативному) со стороны внешней среды, предусматривает проведение следующих действий [2, 7–9, 11].

1. Содержательное описание СЭС и постановка задачи на СИМ, определение исходных данных, результатов СИМ и критериев для их оценки.

2. Проведение комплексного исследования СЭС: сбор и обработка всей доступной информации о ней для проведения СИМ.

3. Идентификация законов распределения исходных данных, расчет статистических оценок параметров этих законов.

4. Математическое моделирование блоков СЭС и СЭС в целом, программирование СИМ-модели, разработка плана компьютерного эксперимента.

5. Имитация процесса функционирования СЭС на СИМ-модели.

6. Статистическая обработка и интерпретация результатов СИМ.

Можно видеть, что идеология СИМ по МДМ гораздо шире идеи использовать технологию метода Монте-Карло (ММК) для исследования и прогнозирования параметров СЭС [7, 8]. Известны два подхода к СИМ: школа Форрестера—Медоуза полагает, что данные пп. 1–3 допустимо задавать в первом приближении, а затем варьировать их при проведении СИМ [11]. Затраты времени на решение исследовательских задач, с учетом параметров современных ЭВМ, не играют особой роли, однако при работе с такой "игрушечной" моделью трудно сохранить объективность принимаемых решений и, напротив, легко получить любой желаемый (в том числе ошибочный) результат [2].

Подход большинства отечественных специалистов предполагает использование в интересах СИМ теории организации и планирования научного эксперимента [3–8]. Достоверность всех фигурирующих при этом данных (исходных, промежуточных, итоговых) тесно связана с их кумулятивностью, под которой будем понимать свойство данных минимального объема адекватно отображать необходимые характеристики исследуемого объекта [9]. Считается, что достичь кумулятивности можно с помощью упомянутого комплексного исследования СЭС, т. е. путем сбора и обработки необходимого объема информации по пп. 1–3.

При необходимости моделировать объекты с малой прецедентной базой (проектируемые, инновационные, разведанные) возникают трудности, для преодоления которых МДМ использует возможности ММК. Решение проблемы кумулятивности при этом сводится к рациональному вербальному моделированию СЭС, в ходе которого важно найти правильное сочетание детерминированных и статистических методов, используемых при проведении СИМ.

На рис. 2 штриховым контуром обозначены этапы, определяющие точность и достоверность исходных данных для проведения СИМ. Поскольку реальная СЭС является основным источником информации о себе самой, определить эти данные (соответствующие им характеристики и параметры СЭС обозначим как группу  $X_1; X_2, \dots, X_N$ ) можно в рамках проведения исследования СЭС, итогом которого являются статистические сведения, относящиеся к указанной группе. После анализа и обработки первичных данных (в виде гистограмм, графиков, числовых оценок параметров и т. п.) определяются типовые законы распределения для  $X_1; X_2, \dots, X_N$  из

числа известных и доступных для разработчиков СУ вероятностных функций (типовых законов), а также проводится оценка их параметров, обычно отличающихся от полученных ранее первичных оценок.

Затем с помощью технологии ММК "разыгрываются" массивы случайных чисел  $X'_1; X'_2, \dots, X'_N$ , соответствующих им  $X_1, X_2, \dots, X_N$ , которые совместно с другими группами исходных данных, формируют состав воздействий  $A_1; A_2, \dots, A_N$  на входы



Рис. 2. Этапы СИМ, влияющие на точность и достоверность исходных данных

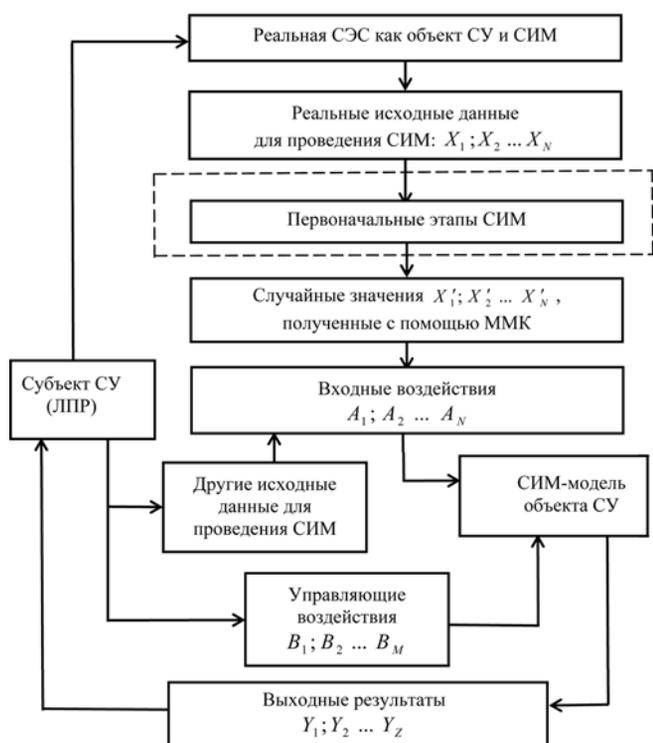


Рис. 3. Схема СУ объектом с применением его СИМ-модели

СИМ-модели. На схеме рис. 3 помимо указанных этапов СИМ и входных воздействий  $A_1; A_2, \dots, A_N$  показаны воздействия  $B_1; B_2, \dots, B_M$ , необходимые для управления СИМ-моделью, а также выходные данные  $Y_1; Y_2, \dots, Y_Z$ , т. е. результаты СИМ, которые субъект СУ (в простейшем случае это ЛПР) использует для управления СЭС по внутреннему контуру (внешний контур на рис. 3 не показан). Отметим, что в адекватной СИМ-модели управляющие воздействия  $B_1; B_2, \dots, B_M$  должны в требуемой мере соответствовать факторам (методам и средствам), которые позволяют управлять реальной СС.

Мы видим, что неопределенности (неточности, неадекватности, нестационарности), присущие исходным данным  $X_1, X_2, \dots, X_N$ , преобразуются в неопределенности входных воздействий  $A_1; A_2, \dots, A_N$  и результатов СИМ  $Y_1; Y_2, \dots, Y_Z$ , т. е. непосредственным образом "участвуют" в формировании качества управленческих решений (см. рис. 3), потенциально ухудшая его. Поэтому уменьшение этих неопределенностей или снижение их влияния на результаты СИМ по МДМ представляют для разработчиков СУ не только теоретический, но и важный практический интерес.

### Применение СИМ для решения прикладных задач

Ограничимся двумя примерами использования СИМ по МДМ для решения задач, связанных с управлением СЭС. Первым примером является управление бизнес-процессом выполнения заказов (удовлетворения заявок) клиента [1, 2]. Заказами при этом считаются любые требования клиента, а исполнителем — любая организация, выполняющая эти требования (абстрактная СЭС, реализующая абстрактную продукцию). В упрощенном виде этот алгоритм применяется для учебных целей, в более сложном — для управления инфокоммуникационной компанией (ИКК). В последнем случае обобщенный алгоритм для прогнозирования процесса оплаты населением услуг крупной региональной ИКК при 34 входных параметрах СИМ-модели включает более 200 укрупненных шагов-действий, объединенных в следующие блоки:

- ввод исходных данных;
- объявление и обнуление переменных СИМ;
- моделирование поступления заявок;
- сравнение момента поступления заявки с окончанием периода моделирования;
- моделирование обслуживания заказов (обработки заявок);
- расчет итоговых показателей и вывод результатов СИМ.

Выходными данными СИМ здесь являются: число включенных и отключенных за период моделирования абонентов; два графика, иллюстрирующие динамику изменения числа абонентов (рис. 4, а) и дебиторской задолженности (рис. 4, б); общая сумма начислений и оплаты по дням; число претен-

зий, процент оплаты по ним и другие показатели. Важными результатами СИМ являются величина упущенной выгоды ИКК от перерывов в предоставлении услуг, а также затраты, связанные с изменением тарификации абонентов при их включении и отключении.

Второй пример — моделирующий алгоритм бизнес-процесса восстановления работоспособности (реанимирования) нефтяных скважин [10], который в качестве СЭС имеет важное практическое значение по трем причинам. Во-первых, его успешное выполнение ведет к значительному экономическому эффекту; во-вторых, он определяет правильность распределения трудовых ресурсов нефтедобывающей компании (НДК); в-третьих, существенно влияет на характер экологического взаимодействия НДК с окружающей средой. Особенностью данной СИМ-модели оказалась неожиданно тесная взаимосвязь "технического" и "организационного" компонентов СЭС, поскольку для рационального распределения трудовых ресурсов НДК необходимо правильно выбирать порядок реанимирования скважин, находить наиболее экономичные методы и средства их восстановления, которые достаточно сложным образом зависят от целого ряда технических параметров объекта и геолого-минералогических характеристик месторождения. Поэтому в работе [10] применение СИМ по МДМ для уменьшения неопределенности сочетается с методами факторного анализа и обработки многочисленных и разнородных данных (статистических и детерминированных), включая: статический и динамический

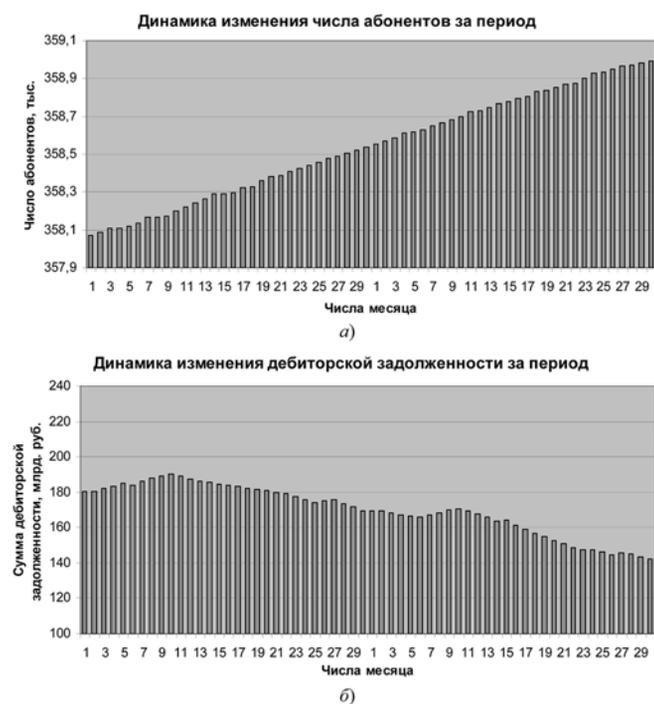


Рис. 4. Прогноз динамики показателей работы ИКК на два месяца: а — число абонентов; б — сумма дебиторской задолженности

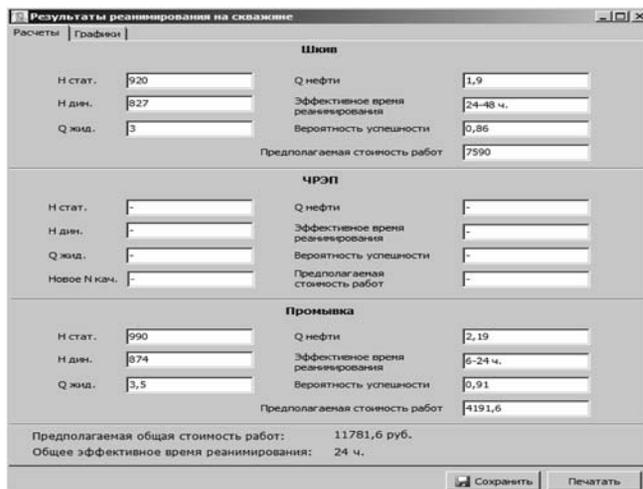


Рис. 5. Результаты СИМ-реанимирования скважины верхнего горизонта

уровни жидкости в каждой скважине; дебиты жидкости и нефти; характеристики способа реанимирования (с помощью частотного регулируемого преобразователя, универсального шкива, промывки химическими реагентами); параметры скважин и т. п.

На рис. 5 представлено окно итогов СИМ по МДМ скважины на верхнем горизонте месторождения, откуда видно, что реанимирование с помощью универсального шкива не приводит ее в работоспособное состояние: статический и динамический уровни жидкости (920 и 827 м) не достигли прежних режимных значений (937 и 851 м); дебит жидкости и нефти в сутки аналогичным образом составляют 3 м<sup>3</sup> и 1,9 т (было 3,1 м<sup>3</sup> и 2 т); ожидаемое время работ 24...48 ч при вероятности успешности 0,86 и стоимости 7590 руб. Напротив, комбинированный вариант "универсальный шкив + + промывка" ведет к требуемому росту статического (до 990 м) и динамического (до 874 м) уровней жидкости; увеличению дебита жидкости и нефти (до 3,5 м<sup>3</sup> и 2,19 т в сутки) с вероятностью 0,91 при стоимости работ 11 781 руб. в течение 6...24 ч.

По итогам СИМ в автоматическом режиме СУ формирует планы-графики для бригад восстановления скважин. Каждый такой план-график содержит информацию о номерах бригады и скважины; варианте реанимирования; новом числе качаний станка-качалки; виде промывочной жидкости (при необходимости промывки); дате и времени начала работ; оптимальной продолжительности восстановления; вероятности успешности; делаются также отметки о принятии к исполнению и дате фактического исполнения (которые заносятся в базу данных НДК после завершения работ на скважине).

#### Статистическое исследование объекта: резервы и новые возможности

Трудность устранения влияния неопределенности исходных данных на результаты моделирования

хорошо известна ЛПР — разработчикам СС и СУ. "Как ни труден отбор надежных данных в физике, гораздо сложнее собрать обширную информацию экономического или социологического характера, состоящую из многочисленных серий однородных данных... В этих обстоятельствах безнадежно добиваться слишком точных определений величин, вступающих в игру. Приписывать неопределенным по самой своей сути величинам какую-то особую точность бесполезно и нечестно, и, каков бы ни был предлог, применение точных формул к этим слишком вольно определяемым величинам есть не что иное, как обман и пустая трата времени" [12].

В то же время: "Многие не признают потенциальной пользы модели, основываясь на том, что у нас нет достаточных данных для моделирования. Они уверены, что первым шагом должен быть широкий сбор статистических сведений. Верно же как раз обратное". И далее: "Мнение о том, что математическая модель не может быть построена, пока не будут полностью известны каждая константа и функциональная зависимость, является недоразумением" [11]. Принимая во внимание эти и другие имеющиеся (в чем-то схожие, в чем-то противоречащие друг другу) точки зрения, МДМ избирает стратегию "стучитесь, и вам откроется" на основании следующих аргументов.

Согласно теории управления СС, в основу принятия решений могут быть положены экспертные, теоретико-вероятностные, вероятностно-статистические и статистические методы и модели (рис. 6). Применительно к проектированию СУ СЭС экспертные методы представляются слишком "слабыми", а статистические — слишком жесткими. Лучшее решение — это СИМ по МДМ, сочетающее достоинства теоретико-вероятностного и статистически-вероятностного подходов, а также технологию ММК [4—5]. Сегодня СИМ является одним из универсальных и эффективных средств исследования СС: математики используют его при проведении компьютерных экспериментов, призванных проверить и подтвердить аналитические выкладки. Прикладные специалисты видят в СИМ средство решения задач, не решаемых другими способами. Системные аналитики применяют СИМ, когда объем знаний об иерархии СС существенно меньше знаний о ее подсистемах и элементах. Менеджеры заинтересованы в управлении бизнес-процессами СЭС с помощью СИМ (см. рис. 3). Поэтому предложение использовать данный метод при проектировании СУ представляется закономерным.

При изучении СЭС в интересах проектирования СУ задача СИМ заключается в исследовании реакции разных вариантов реализации СЭС на комбинации разных (позитивных и негативных) воздействий на нее. Соответственно, управление СЭС предусматривает разработку СУ, которая в заданной (максимально возможной, минимально не-



Рис. 6. Методы и модели теории СС, применимые при разработке СУ



Рис. 7. К определению принципов моделирования и проектирования СУ СЭС

обходимой) мере призвана усилить последствия позитивных воздействий (за счет положительных обратных связей) и препятствовать негативным воздействиям за счет отрицательных обратных связей. Целью СИМ при этом является создание и применение моделей СЭС и СУ, которые отражают их кумулятивные (минимальные по объему и максимальные по содержательности) характеристики, необходимые и достаточные для проектирования СУ, что в схематичном виде иллюстрирует рис. 7.

Отметим, что схема рис. 7 отражает идеологию любого моделирования СС: мысленного, вербального, физического, математического, компьютерного имитационного и т. д. [9]. Во всех этих случаях после выполнения ряда первоначальных этапов (по аналогии со схемой рис. 2) осуществляется переход из реальной среды в виртуальную, где исследуются свойства моделей СС и СУ с тем, чтобы полученные результаты затем вернуть из виртуальной среды в реальную.

Эффективность и практическая значимость такого возвращения зависят от двух причин: во-первых, они обусловлены точностью и адекватностью используемых моделей, во-вторых, соответствием их параметров характеристикам реальных объектов. В этом смысле кумулятивность СИМ-модели означает отсутствие у нее управляемых параметров, которые не имеют себе аналогов на реальном объекте, что сразу обнаруживает целый ряд недостатков моделей, обычно рекомендуемых при проектировании СС.

Школа Форрестера—Медоуза не считает препятствием для проведения СИМ отсутствие надежных и достоверных исходных данных, но ее подход не обеспечивает возможность управления СЭС, поскольку возникающие неопределенности могут быть слишком большими. В то же время проведение трудоемких исследований на объекте и преодоление сложностей, возникающих при моделирова-

нии конкретных СЭС, также не оправдывают себя, если итог сводится к выбору из двух-трех типовых моделей, которым примерно в равной мере не противоречат (например, в соответствии с критерием Пирсона) полученные экспериментальные данные [2, 9]. Поэтому МДМ считает возможным не исключать, но максимально упростить и автоматизировать первоначальные этапы СИМ (см. рис. 3):

- во-первых, так как на рынке сегодня присутствуют программные продукты, предназначенные для обработки статистических данных с идентификацией соответствующих им законов распределения и оценкой параметров этих законов [13];
- во-вторых, потому что для обширного класса устойчивых (робастных) СЭС результаты СИМ не должны (и не будут на практике) в критической мере зависеть от правильности действий ЛПР на данном этапе;
- в-третьих, нужно иметь в виду, что неопределенность исходных данных означает недостаток или неполноту верифицированных знаний о СЭС как объекте СИМ, но не их отсутствие, а тем более недостаток или отсутствие соответствующих аксиологических знаний [14], что в рамках МДМ дает новые возможности для развития СИМ.

### Заключение

Дальнейшее повышение эффективности использования СИМ по МДМ для управления СЭС может быть связано с применением теории риска [15], а также разновидностей неоклассической теории вероятностей (ТВ) — теории шансов и ожидаемой (объективной и субъективной) полезности. Введение в анализ вместо объективной частотной вероятности значений субъективной вероятности, отражающих опыт и знания ЛПР, выражающих их предпочтение и познавательную активность, еще более сближает теоретико-вероятностные и экспертные методы на рис. 6.

Важно также, что СИМ по МДМ для достаточно большого класса СС (как чисто каузальных, так и обладающих стохастическими свойствами) легко и просто сочетается с теорией риска [16]. В рамках концепции "риск как неопределенность" его значение считается объективно существующей случайной величиной, связанной с вероятностным распределением возможных исходов некоторой операции (процедуры, процесса), что характерно для традиционной ТВ. При концепции "риск как возможность", которая использует теоретико-вероятностные и экспертные модели (см. рис. 6), теория риска сближается с теорией ожидаемой полезности и ТВ шансов.

Существенны два обстоятельства. Во-первых, все эти исследовательские действия происходят в виртуальной среде и относятся к СИМ-моделям СС (см. рис. 7), а не связаны с изучением объекта в реальной среде, результатом которого могут быть упо-

мянутые "обман и пустая трата времени", хотя для достижения цели СИМ этого достаточно. Во-вторых, речь здесь идет не об аппроксимации неполных, неточных, недостоверных и неадекватных данных, полученных путем предварительного изучения СЭС, а именно о моделировании объекта, в качестве своеобразной характеристики которого выступает неопределенность знаний о нем.

С точки зрения теории СС здесь путаницы нет, так как неопределенность знаний о СЭС, являющейся объектом СИМ, сама может считаться виртуальной СС и быть объектом СИМ. Развитие и совершенствование методов и средств СИМ по МДМ, направленное на снижение неопределенности исходных, промежуточных и выходных данных, имеет важное практическое значение, поскольку способствует повышению эффективности действий ЛПР, связанных с управлением СЭС и других СС организационно-технического типа.

### Список литературы

1. Димов Э. М., Маслов О. Н., Скворцов А. Б. Новые информационные технологии: подготовка кадров и обучение персонала. Ч. 1. Реинжиниринг и управление бизнес-процессами в инфокоммуникациях. М.: ИРИАС, 2006. 386 с.
2. Димов Э. М., Маслов О. Н., Пчеляков С. Н., Скворцов А. Б. Новые информационные технологии: подготовка кадров и обучение персонала. Ч. 2. Имитационное моделирование и управление бизнес-процессами в инфокоммуникациях. Самара: Изд-во СНЦ РАН, 2008. 350 с.
3. Голенко Д. И., Димов Э. М., Осокин В. В. Вопросы оптимального планирования в системах управления дискретно-непрерывным производством // Опыт применения прикладных методов математики и вычислительной техники в народном хозяйстве. М.: Статистика, 1974. С. 15—32.
4. Димов Э. М. Оптимальное управление комплексным производством дискретно-непрерывного типа на основе имитационного моделирования // Опыт применения прикладных методов математики и вычислительной техники в народном хозяйстве. М.: Статистика, 1975. С. 58—73.
5. Димов Э. М. Имитационное моделирование и оптимизация управления в сложных производственных системах. Саратов: Изд-во СГУ, 1983. 165 с.
6. Моисеев Н. Н. Элементы теории оптимальных систем. М.: Наука, 1975. 526 с.
7. Димов Э. М., Маслов О. Н. О развитии математических принципов метода имитационного моделирования // Инфокоммуникационные технологии. 2003. Т. 1, № 2. С. 5—11.
8. Димов Э. М., Маслов О. Н. О точности и адекватности метода статистического имитационного моделирования // Инфокоммуникационные технологии. 2007. Т. 5, № 1. С. 60—67.
9. Павловский Ю. Н. Имитационные модели и системы. М.: ФАЗИС, ВЦ РАН, 2000. 134 с.
10. Патент RU 2 469 370 С1. Способ реанимирования нефтяных скважин // Димов Э. М., Маслов О. Н., Халимов Р. Р. и др. Заявл. 16.05.2011, опубл. 10.12.2012, бюлл. № 34.
11. Форрестер Дж. Основы кибернетики предприятия (индустриальная динамика). Пер. с англ. М.: Прогресс, 1971. 340 с.
12. Винер Н. Творец и робот: Пер. с англ. М.: Прогресс, 1996. 104 с.
13. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: для профессионалов. СПб.: Питер, 2003. 688 с.
14. Виттик В. А. Организация сложных систем. Самара: Изд-во СНЦ РАН, 2010. 66 с.
15. Королев В. Ю., Бенин В. Е., Шоргин С. Я. Математические основы теории риска. М.: Физматлит, 2007. 544 с.
16. Маслов О. Н. Безопасность корпорации: моделирование и прогнозирование внутренних угроз методом риска. Самара: Изд-во ПГУТИ-АЭРОПРИНТ, 2013. 170 с.

УДК 621.3.049.771.14

**С. В. Гаврилов**, д-р техн. наук, зав. отделом, e-mail: sergey\_g@ippm.ru,  
**Д. И. Рыжова**, мл. науч. сотр., e-mail: ryzhova\_d@ippm.ru,  
**А. Л. Стемпковский**, д-р техн. наук, проф., академик РАН, директор,  
Институт проблем проектирования в микроэлектронике РАН, г. Москва, Зеленоград

## Проблема анализа пикового тока при проектировании сверхбольших интегральных схем на логическом уровне и современные методы ее решения

*Проанализированы ключевые проблемы проектирования сверхбольших интегральных схем, для которых необходима оценка пикового тока, и известные подходы к их решению. Рассмотрены преимущества и недостатки различных методов анализа пикового тока. Разработаны методы оценки пикового тока КМОП-схем на логическом уровне моделирования, обеспечивающие повышение точности по сравнению с известными методами.*

**Ключевые слова:** статический временной анализ, сложно-функциональный блок, логические корреляции, анализ пикового тока

S. V. Gavrilov, D. I. Ryzhova, A. L. Stempkovsky

## Problem of Peak Current Analysis at the Logic Level of VLSI Design and its Modern Solutions

*Key problems of very large scale integrated circuits (VLSI) design for which peak current estimation is required and known approaches to their solution are analyzed. Advantages and disadvantages of different methods of peak current analysis are considered. Methods of peak current estimation for CMOS circuits at the logic simulation level which provide improved accuracy compared with the known methods are developed.*

**Keywords:** static timing analysis, intellectual property block, logic correlations, analysis of peak current

### Введение

Важной особенностью проектирования современных сверхбольших интегральных схем (СБИС) является необходимость разработки и использования моделей, адекватно отражающих поведение субмикронных и нанометровых элементов на различных уровнях описания. Для диапазона 10...100 нм значительно изменяются многие свойства и характеристики твердых тел, и приходится учитывать в моделях многие физические процессы и явления, которые были несущественными при больших размерах. Для нанометрового диапазона размеров элементов начинают проявляться в полной мере квантовые эффекты, что требует разработки принципиально новых моделей и методов проектирования.

Ключевыми проблемами, для которых необходим расчет пикового тока, являются эффект IR-drop [1, 2] — падение напряжения в цепях земли/питания — и расчет ширины шин питания при проектировании схем. С уменьшением технологических норм эти проблемы становятся ограничивающими факторами в разработке СБИС.

Эффект IR-drop приводит к возникновению напряжения в проводнике в результате его электрического сопротивления (рис. 1). Это напряжение пропорционально току, который течет через проводник ( $V_{drop} = IR$ ), и приводит к падению напряжения.

Топологические размеры элементов СБИС уменьшаются с каждым годом, и, соответственно, для них определяются свои рабочие напряжения питания и запасы помехоустойчивости. Тем не ме-

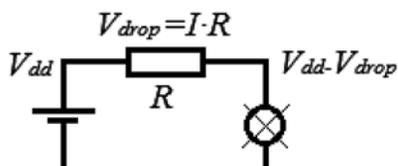


Рис. 1. Иллюстрация к эффекту IR-drop

нее, для многих устройств данные характеристики не масштабируются, и в случае, например, энергоёмких устройств, работающих на низком напряжении, в них может появиться значительный мгновенный ток. Для узких шин питания/земли эти кратковременные "всплески" тока могут вызвать перенапряжение и выход схемы из строя. Кроме того, низкое напряжение питания свидетельствует об уменьшении запаса помехоустойчивости, необходимого для корректной работы транзисторов на кристалле. Для того чтобы избежать логических ошибок, необходимо проектировать схему с учетом падения напряжения и уменьшения запаса помехоустойчивости. Поэтому для обеспечения работоспособности СБИС необходимо учитывать максимальный мгновенный ток, протекающий в схеме.

Анализ существующих подходов к оценке пикового тока показал, что они не предоставляют достаточно точного решения поставленной задачи. Возникает необходимость в разработке методов оценки пикового тока на логическом уровне моделирования, обеспечивающих повышение точности по сравнению с известными методами.

Дальнейшее содержание работы следующее: в разделе 1 представлен анализ существующих методов оценки пикового тока, в разделе 2 приведено описание нового подхода к решению задачи расчета пикового тока с учетом одновременного переключения входов.

## 1. Анализ существующих методов оценки пикового тока

### 1.1. Модели задержек и токов на вентиляльном уровне

Одной из самых распространенных моделей в современных САПР является NLDM-модель ("Non-linear delay model"). Она представляет нелинейные зависимости задержек  $D_{out}$  выходных сигналов от длительности фронта  $S_{inp}$  входного сигнала и емкости  $C_{out}$  нагрузки в форме двумерных таблиц

$$D_{out}(S_{inp}^k, C_{out}^l), k \in [1: N_s], l \in [1: N_c].$$

Для проведения анализа схем, содержащих RC-модели межсоединений, с более высокой точностью, чем в NLDM, в компании Synopsys была разработана другая модель, получившая название "Composite current source" (CCS). Модель состоит из двух основных частей — CCS-модель драйвера и CCS-модель приемника.

CCS-модель драйвера представляет собой источник тока, который зависит от времени  $t$  и от тех же параметров, что и в NLDM —  $S_{inp}$  и  $C_{out}$ :

$$I_{out} = F(t, S_{inp}, C_{out}).$$

При фиксированных значениях  $S_{inp}$  и  $C_{out}$  функция  $F(t, S_{inp}, C_{out})$  представляет собой ток через нагрузочную емкость  $C_{out}$  при подаче на вход вентиля сигнала с длительностью фронта  $S_{inp}$ . Ток  $I_{out}(t)$  имеет колоколообразную форму. В CCS-модели функция  $I_{out}(t)$  запоминается в виде набора дискретных точек  $\{t^k, I_{out}^k\}$  на отрезке  $[t_0, t_1]$ . Фиксируется значение времени  $T_{ref}$ , когда входной сигнал  $V_{inp}(t)$  пересекает пороговый уровень  $0,5V_{dd}$ , где  $V_{dd}$  — напряжение питания.

В CCS-модели приемника входы так же, как и в NLDM, представляются емкостями, но в CCS предусматривается зависимость этих емкостей от  $S_{inp}$  и  $C_{out}$ . Таблицы эффективных входных емкостей связываются с каждой дугой логического графа  $arg_{ij}$ :

$$C_{-1}_{inp}^{(r,f)}(S_{inp}, C_{out}), C_{-2}_{inp}^{(r,f)}(S_{inp}, C_{out}),$$

где  $C_{-1}$  — это таблица значений эффективных входных емкостей, рассчитанных для половины фронта входного сигнала, от начала переключения до  $T_{ref}$ ;  $C_{-2}$  — это таблица значений этих же емкостей, вычисленных для второй половины фронта входного сигнала, от  $T_{ref}$  до конца переключения.

Объем данных для CCS-модели намного превышает число данных для NLDM, но их получение требует такое же число расчетов электрической схемы. Точность данной модели приближается к электрической модели вентиля.

В ряде систем, в том числе в продуктах компании Cadence, применяется сходная с CCS-моделью модель, называемая "Effective current source model" (ECSM). ECSM-модель имеет два существенных отличия от CCS. Во-первых, вместо функций  $I_{out} = F(t, S_{inp}, C_{out})$ , в ECSM записываются табличные значения функций  $V_{out} = G(t, S_{inp}, C_{out})$ , где  $V_{out}$  — напряжение на выходе вентиля. Значения функции  $V_{out}(t)$  хранятся в виде одномерных таблиц для отрезка времени  $[t_0, t_1]$ , где точки  $t_0, t_1$  определяются условием

$$V_{out}(t) \in [0 + \varepsilon, V_{dd} - \varepsilon],$$

где  $\varepsilon$  — заданная константа. Помимо этого значения  $V_{out}(t)$  умножаются на  $\frac{1}{V_{dd}}$ , т. е. приводятся к интервалу (0, 1). Во-вторых, так же, как и в CCS, моделями входов являются таблицы емкостей  $C_{inp}^{(r,f)}(S_{inp}, C_{out})$ , но они не делятся на две таблицы

$C_1$ ,  $C_2$ . Индексы  $r/f$  используются для обозначения направления переключения сигнала  $r(rise)$  из 0 в 1,  $f/fall$  из 1 в 0.

Применение описанных выше моделей задержек позволило значительно увеличить точность логического моделирования субмикронных цифровых КМОП-схем, но остались нерешенными следующие проблемы:

- такие модели не учитывают одновременные переключения вентиляей;
- с помощью таких моделей нельзя получить зависимость тока в цепи питания от времени  $I_{V_{dd}}(t)$ ;
- такие модели не учитывают влияние активных сопротивлений цепей питания (IR-drop эффект).

### **1.2. Средства моделирования тестовых последовательностей**

В настоящее время при реализации проектов цифровых схем актуальной является проблема тестирования СБИС на языках описания аппаратуры с использованием современных САПР. Для обеспечения процесса тестирования цифровых СБИС недостаточно просто синтезировать тестопригодную схему. Необходимо создать набор тестов, гарантирующих качественную проверку функционирования разрабатываемого устройства. Применение средств автоматической генерации тестов значительно облегчает решение этой задачи.

На стадии разработки проекта необходимо заложить в него возможности тестирования для выявления технологических дефектов. Для этого служат специальные средства САПР для логического синтеза тестопригодных проектов (Design-for-test). Далее происходит функциональная верификация проекта на соответствие выполняемых функций спецификации. Итогом функциональной верификации являются функциональные векторы. Они представляют собой последовательность сигналов, снимаемых через определенный интервал времени с входов и выходов кристалла (временной срез). Далее необходимо сгенерировать специальные тестовые векторы для определения возможных дефектов кристалла. Этим занимаются специальные средства САПР, называемые ATPG (Automatic Test Pattern Generation). Следующими этапами являются размещение элементов, трассировка и уже непосредственное изготовление кристалла дизайн-фабрикой.

Одними из основных производителей САПР, предоставляющих средства подготовки тестопригодных проектов, верификации и создания тестовых векторов, являются Cadence и Synopsys. Средства UltraSim [3] и Encounter [4] являются современными инструментами для моделирования тестовых последовательностей. В компании Synopsys к средствам синтеза тестопригодных проектов можно отнести Design Compiler, DFT Compiler, BSD Compiler. Средство ATPG — TetraMAX ATPG, симуляторы — Nanosim [5], CustomSim [2].

Моделирование тестовых последовательностей позволяет получить нижнюю оценку пикового тока. Значение тока рассчитывается симуляторами как разность напряжений двух терминалов резистора, деленная на его сопротивление. Полное моделирование схемы — процесс затратный и дорогостоящий, а в случае больших схем — невозможный из-за большой размерности задачи.

### **1.3. Методы генерации тестовых последовательностей для оценки пикового тока**

Обширные исследования по оценке тока в схемах для детерминированных моделей входов были проведены в работе [6]. Один из способов найти пиковый ток — это исчерпывающий поиск по всем возможным комбинациям входных последовательностей. Однако эти методы можно применять только к небольшой схеме, имеющей несколько входов. С увеличением числа входов схемы число входных последовательностей растет экспоненциально, поэтому данный метод не практичен для больших схем.

Чоудхури и др. [7] обратились к проблеме оценки максимального тока для больших схем с большим числом входов. В предлагаемых ими методах большая схема делится на более мелкие логические блоки. Тогда существует метод поиска, например, эвристический метод для нахождения максимального тока для каждого логического блока. Сумма максимальных переходных токов для каждого блока представляет собой оценку максимального тока для всей схемы. В итоге оценка получается излишне пессимистичной. Кроме того, из-за большого размера схемы поиск решения осуществляется очень медленно.

Подход, описанный в работе [8], предлагает алгоритм генерации тестовых последовательностей с учетом единичной задержки на вентилях схемы для оценки максимального мгновенного тока через шины питания. В этой работе впервые была представлена токовая модель с учетом одновременного переключения входов схемы. Временная ось разбивается на интервалы (временные слоты), в которых возможно переключение вентиляей, в каждом слоте рассчитывается ток переключения в цепи питания. В качестве пикового тока принимается максимальное значение из рассчитанных в каждом временном слоте. В результате скорость моделирования увеличивается по сравнению с полным перебором всех возможных входных комбинаций [6], и оценка пикового тока оказывается не такой завышенной, как в случае суммирования всех максимальных токов [7].

Данный подход имеет следующие недостатки:

- используется упрощенная модель задержки, равная единичной задержке на всех вентилях;
- требуются большие временные и аппаратные затраты для расчета больших схем.

#### 1.4. Методы верхних оценок пикового тока

В работе [9] верхняя граница максимального тока в шинах питания/земли оценивается с помощью распространения входных сигналов во временных интервалах через нормированные цепочки вентиляей. Предлагаемый алгоритм, называемый "iMax", рассчитывает максимальный ток сигнала в интервалах времени с помощью независимой модели. Таким образом, этот подход относится к статическому методу расчета тока, так как моделирование больших наборов входных воздействий для крупных схем является дорогостоящим. При таком подходе используются следующие допущения:

- все входы независимые;
- все первичные входы переключаются в момент времени  $t = 0$ ;
- задержки вентиляей фиксированы и известны заранее;
- различные параметры тока, такие как пик, длительность и фиксируемое время, рассчитываются в этапе предварительной обработки из схемы на уровне параметров, рассматриваемых и связанных с их входами и выходами вентиляей;
- на стыке временных интервалов входы не меняются и, следовательно, нет соответствующего интервала неопределенности на выходе схемы.

Оценка пикового тока должна учитывать и число переключений, и число узлов с большими выходными емкостями во время поиска максимума. Нахождение максимума само по себе недостаточно для обеспечения жесткой оценки пиковой мощности, так как необходимо учитывать также одновременное переключение нескольких вентиляей.

### 2. Разработка методов оценки пикового тока КМОП-схем на логическом уровне с учетом одновременного переключения

#### 2.1. Модели "ток—задержка"

Наиболее простые модели тока, известные сегодня из публикаций, — это модели прямоугольника [7], треугольника [10], трапеции [11], они не отражают характер реальной кривой тока, но учитывают значение максимального тока.

Для повышения точности моделей в качестве прототипа для разрабатываемого подхода были выбраны CCS-модели и ECSM-модели, которые используются для других целей. Они сохраняют табличные значения функций  $I_{out}$  и  $V_{out}$  от времени соответственно.

В рамках исследований были построены зависимости пикового тока от входных фронтов и смещения между ними при одновременном переключении нескольких входов (рис. 2). Очевидно, что поведение тока в конкретной схеме зависит от значения смещения между фронтами. При достаточно больших значениях модуля смещения эта зависимость исчезает, и оценка тока сводится к случаю

переключения по одному входу. Значение пикового тока достигает максимума при малых значениях смещения, не всегда равных нулю.

По результатам моделирования было принято решение о разработке более точной модели на основе метода контрольных точек. В этой модели значения тока во времени сохраняются по выбранным уровням входных и выходных напряжений:

$$I_0 = \{(t_i, I_i): V_0(t_i) = OL_i \cdot V_{dd}\},$$

$$I_1 = \{(t_i, I_i): V_1(t_i) = IL_i \cdot V_{dd}\},$$

$$I_2 = \{(t_i, I_i): V_2(t_i) = IL_i \cdot V_{dd}\},$$

где  $IL = \{IL_i\} = \{0,0 \ 0,1 \ 0,2 \ 0,5 \ 0,8 \ 1,0\}$  — уровни входных сигналов,  $OL = \{OL_i\} = \{0,01 \ 0,1 \ 0,2 \ 0,5 \ 0,8 \ 0,9 \ 0,99 \ 0,999\}$  — уровни выходного сигнала.

Было показано, что такой подход обеспечивает достаточно точное соответствие кривой тока, полученной в результате схемотехнического моделирования. На рис. 3 результаты интерполяции на логическом уровне моделирования показаны сплошными линиями, соответствующие результаты схемотехнического моделирования — точками.

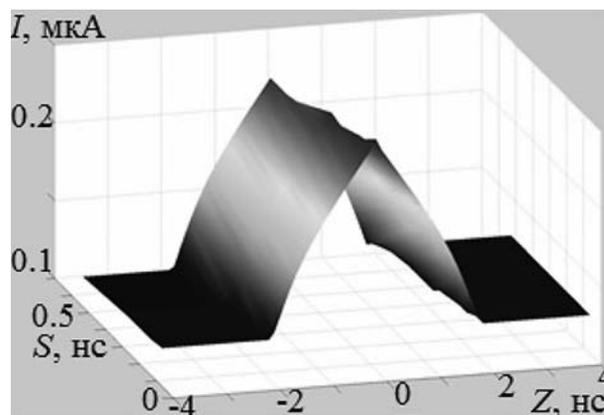


Рис. 2. Зависимость пикового тока  $I$  от фронта  $S$  первого входного сигнала и смещения  $Z$  при фиксированном фронте второго входного сигнала

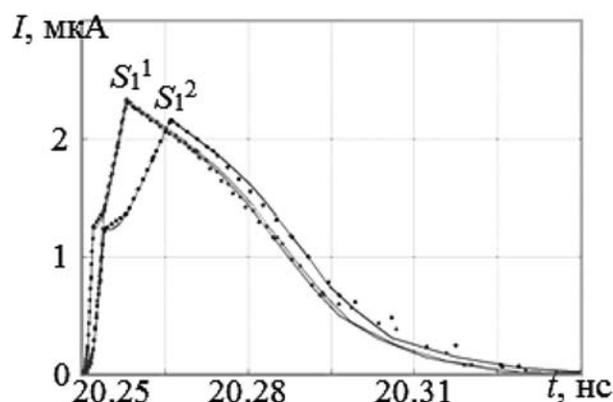


Рис. 3. Зависимость пикового тока  $I$  от времени  $t$  при фиксированных значениях смещения входного фронта  $S_1^1, S_1^2$

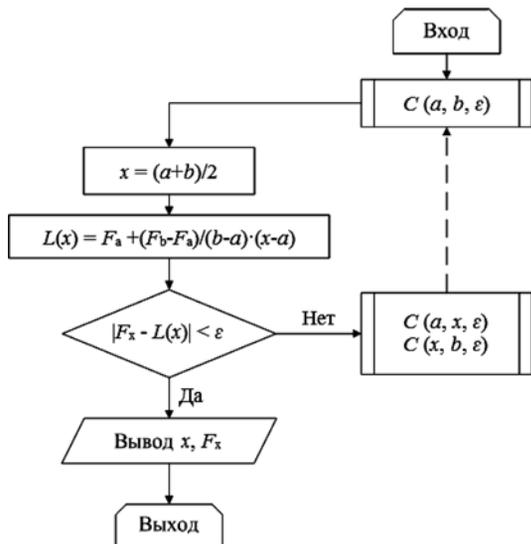


Рис. 4. Алгоритм характеристики тока в контрольных точках на основе двоичного деления по параметру смещения входных фронтов

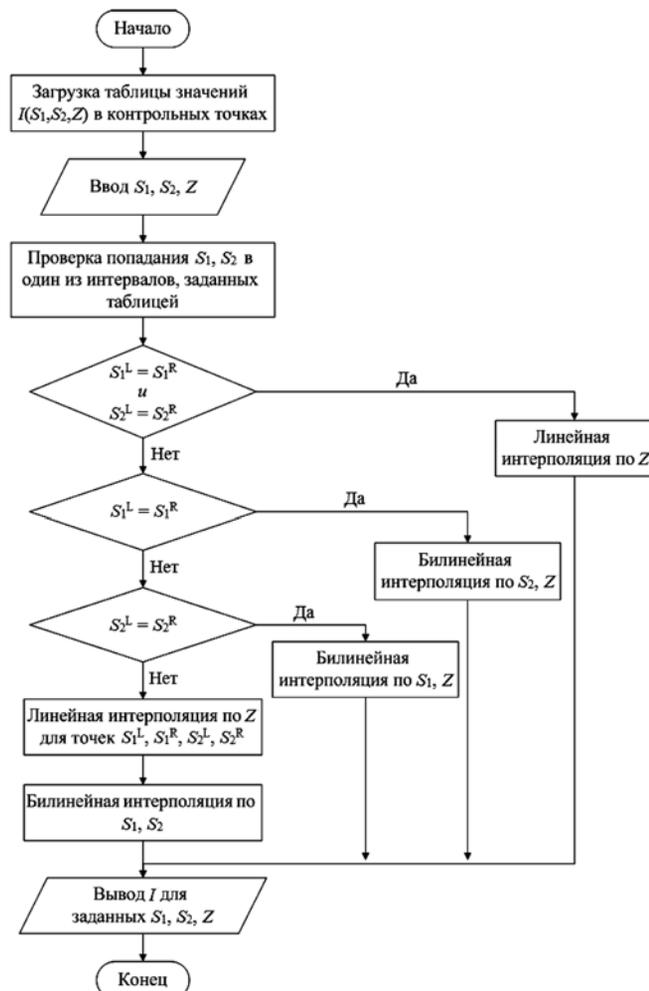


Рис. 5. Алгоритм трилинейной интерполяции токов в контрольных точках

## 2.2. Алгоритм характеристики пикового тока

Для поиска глобального максимума функции тока был использован метод дихотомии. Из полученных в результате применения метода контрольных точек значений тока, фронтов и смещений была составлена таблица характеристики, по которой в дальнейшем можно делать интерполяцию по смещениям между входными фронтами. Алгоритм характеристики тока в контрольных точках на основе дихотомического деления по параметру смещения входных фронтов представлен на рис. 4. Процедура  $C(a, b, \epsilon)$  обеспечивает характеристику на заданном интервале времени  $[a, b]$  с заданной точностью тока  $\epsilon$ . Функция  $L(x)$  применяется для билинейной интерполяции тока в момент времени  $x$  с использованием результатов точного моделирования  $F_a, F_b$  на краях отрезка.

## 2.3. Алгоритм трилинейной интерполяции пикового тока

В классическом подходе на основе билинейной интерполяции (NLDM, CCS, ECSM) каждому значению фронта соответствовало бы одно значение смещения, в нашем случае каждому значению фронта соответствует четыре значения смещения, поэтому необходим иной подход к определению токов в контрольных точках на основе трилинейной интерполяции.

Блок-схема предложенного алгоритма представлена на рис. 5.

Проведенные численные эксперименты по интерполяции пикового тока в контрольных точках показали, что предлагаемый подход обеспечивает эффективный расчет пикового тока на уровне логического моделирования с точностью в пределах 7% по сравнению со схмотехническим моделированием. При этом скорость моделирования возрастает в десятки раз.

## 2.4. Численные эксперименты

В рамках данной работы была выполнена серия численных экспериментов по оценке пикового тока в схемах ISCAS-85 [12] (рис. 6, 7). Результаты метода контрольных точек  $I_{\max}^{CP}$  сравнивались с результатами точного схмотехнического моделирования  $I_{\max}^{Spice}$  по формуле:

$$\Delta = \frac{I_{\max}^{CP} - I_{\max}^{Spice}}{I_{\max}^{Spice}}$$

График  $I_{\max}^{CP}$  показан на рисунках сплошной линией, график  $I_{\max}^{Spice}$  — точками.

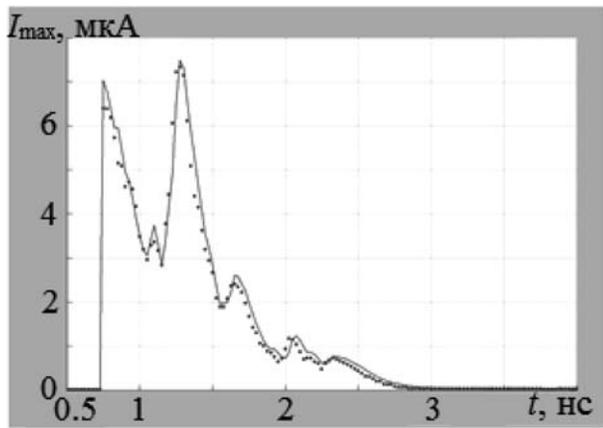


Рис. 6. Сравнение результатов метода контрольных точек и точного моделирования для схемы S432

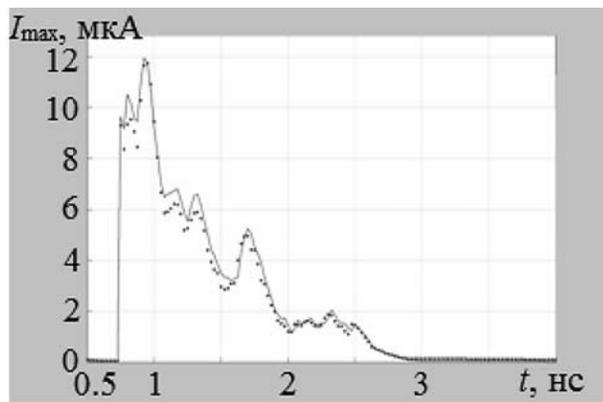


Рис. 7. Сравнение результатов метода контрольных точек и точного моделирования для схемы S1355

Сравнение результатов метода контрольных точек со схемотехническим моделированием показало, что ошибка при нахождении максимального значения тока для больших схем не превышает 7%.

Модель дает точную оценку пикового тока даже в случае появления неполных переключений (glitches) на выходе вентиля. Однако размерность задачи при этом возрастает, и характеристика схем в таких случаях делается "на лету" ("on the fly"). На основе численных экспериментов было показано, что хотя неполные переключения и оказывают существенное влияние, они составляют довольно малый процент всех переключений.

### Закключение

На основе исследования существующих методов оценки пикового тока разработана модель тока на основе метода контрольных точек, которая обеспе-

чивает достаточно точное соответствие кривой тока, полученной в результате схемотехнического моделирования.

Для повышения точности оценки пикового тока на логическом уровне предложен алгоритм характеристики тока потребления вентилях в контрольных точках на основе дихотомического деления по параметру смещения входных фронтов.

Предложен алгоритм построения кривых зависимости тока от времени на основе трилинейной интерполяции токов в контрольных точках.

На основе численных экспериментов по интерполяции пикового тока в контрольных точках показано, что предлагаемый подход обеспечивает эффективный расчет пикового тока на уровне логического моделирования с точностью в пределах 7% по сравнению со схемотехническим моделированием при существенном ускорении на уровне логического моделирования.

*Работа выполнена при поддержке РФФИ (проект № 12-07-00112).*

### Список литературы

1. Selcuk K., Friedman E. Efficient algorithms for fast IR drop analysis exploiting locality // INTEGRATION. 2012. N 45. P. 149–161.
2. Geden B. Understand and Avoid Electromigration (EM) & IR-drop in Custom IP Blocks // Synopsys Webinars. 2011.
3. Jiang W. An Effective EM/IR-drop Flow with UltraSim // CDNLive Silicon Valley. 2007.
4. Encounter Power System. Unified power analysis for faster design optimization and signoff. URL: <http://www.cadence.com/products/di/eps/pages/default.aspx>
5. By Geoffrey Ying. Synopsys, Inc., Mountain View, CA. NanoSim: A Next-Generation Solution for SoC Integration Verification. URL: <http://www.synopsys.com/Tools/Verification/Pages/socintegration.aspx>
6. Hsiao M. S. Peak Power Estimation Using Genetic Spot Optimization for Large VLSI Circuit // Design, Automation and Test in Europe Conference. 1999. P. 175–179.
7. Chowdhury S., Barkatullah J. S. Estimation of maximum currents in MOS IC logic Circuits // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 1990. P. 642–654.
8. Ganeshpure K., Sanyal A., Kundu S. A Pattern Generation Technique for Maximizing Switching Supply Currents // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2012. P. 986–998.
9. Kriplani H., Najm F., Hajj I. Pattern Independent Maximum Current Estimation in Power and Ground Buses of CMOS VLSI Circuits: Algorithms, Signal Correlations and Their Resolution // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 1995. P. 998–1012.
10. Jiang Y. M. Estimation for Maximum Instantaneous Current Through Supply Lines for CMOS Circuits // IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2000. P. 61–73.
11. Sudhakar B. Estimation of maximum current envelope for power bus analysis and design // International symposium on Physical design. 1998.
12. Hansen M., Yalcin H., Hayes J. P. "Unveiling the ISCAS-85 Benchmarks: A Case Study in Reverse Engineering," IEEE Design and Test, vol. 16, N 3, P. 72–80, July–Sept. 1999.

# ГЕОИНФОРМАЦИОННЫЕ СИСТЕМЫ GEOINFORMATION SYSTEMS

УДК 004.6; 528; 004.8

**В. Я. Цветков**, д-р техн. наук, проф., советник,  
Московский государственный технический университет радиотехники, электроники и автоматики  
(МГТУ МИРЭА), e-mail: cvj2@mail.ru

## Ситуационное моделирование в геоинформатике

*Описано ситуационное моделирование в геоинформатике. Его появление стало возможным благодаря накоплению опыта в области моделирования и росту мощности вычислительных ресурсов. Ситуационное моделирование связано с усложнением процесса моделирования. Оно включает раздельное моделирование: ситуации, объекта моделирования и связей между объектом и ситуацией. Ситуационное моделирование позволяет проводить анализ сложных ситуаций и решать задачи, которые обычное моделирование не решает.*

**Ключевые слова:** геоинформатика, геоинформационные технологии, геоинформационное моделирование, ситуационное моделирование, пространственная информация, информационные модели

V. Ya. Tsvetkov

## Situational Modeling in Geoinformatics

*This article describes the situational modeling in geoinformatics. His appearance was made possible by the accumulation of experience in the field of modeling and due to increased power of computing resources. Situational modeling due to the complexity of the modeling process. It includes the separate modeling: situations, object modeling and simulation relations between the subject and situation. Situational modeling allows the analysis of complex situations and solve problems that conventional simulation does not solve.*

**Keywords:** geoinformatics, geoinformation technology, geoinformational modeling, situational modeling, spatial information, information model

### Введение

Геоинформационное моделирование как метод научного познания прошло достаточный период развития с момента становления геоинформатики как науки [1, 2]. По мере роста сложности задач, решаемых методами геоинформатики, развития геоинформационных моделей, совершенствования программного обеспечения, появления новых технологий сбора информации совершенствовались и развивались методы геоинформационного моделирования. В настоящее время возник новый вид моделирования — ситуационное моделирование в геоинформатике, которое требует детального рассмотрения.

### Виды геоинформационного моделирования

Геоинформационное моделирование [3] как обобщение моделирования в геоинформатике делится

на две группы: статическое (стационарное) и ситуационное (динамическое). Статическое моделирование описывает различные статические ситуации. Ситуационное моделирование применяется для анализа динамики ситуаций.

Статическое моделирование в качестве объекта исследования рассматривает *один объект*, будь то объект, явление или ситуация.

Геоинформационное ситуационное моделирование в качестве объекта исследований рассматривает *комплекс качественно разных объектов*. Оно включает моделирование как минимум трех компонентов взаимозависимых сложных систем: объекта исследования; среды вокруг объекта; системы отношений и вытекающих из них связей "объект — среда". Поэтому для понимания специфики ситуационного моделирования первоначально необходимо рассмотреть пространственные отношения.

## Пространственные отношения как основа геоинформационного ситуационного моделирования

Основой геоинформационного статического моделирования [3] являются, в первую очередь, гео-данные и во вторую — пространственные отношения. Многие виды статического геоинформационного моделирования практически не используют пространственные отношения. Эти отношения достаточным образом отражены в гео-данных.

Проводя аналогию с геометрией, можно отметить, что пространственные отношения играют роль глобальной геометрической системы координат, в которой задается локальная система координат (ситуация), по отношению к которой, в свою очередь, определяется объект моделирования. Таким образом, одна из основных функций пространственных отношений состоит в привязке объекта и ситуации вокруг него к некоей глобальной системе, в которой эта ситуация развивается [4].

При ситуационном моделировании пространственные отношения играют доминирующую роль как инструмент описания именно ситуаций, в которой находится объект моделирования или группа объектов. Пространственные отношения задают описание начальной ситуации и определяют динамику ее развития. Связи "объект — среда", которые были упомянуты выше, в геоинформатике задают именно пространственные отношения. Следует подчеркнуть, что пространственные отношения не эквивалентны связям, а только задают условия для их возникновения, существования и развития.

Особенность ситуационного моделирования в том, что во внимание принимается объект моделирования и окружающая его микроситуация или микросреда. Совокупность развития ситуаций образует сценарий. Возможны четыре типа развития простых ситуаций (сценариев), они приведены в табл. 1.

Сценарий 1 соответствует обычному статическому геоинформационному моделированию. Остальные сценарии 2...4 приводят к ситуационному моделированию. Динамику среды задают пространственные отношения. Они же определяют характер взаимодействия объекта и среды.

В более сложном случае ситуационное моделирование включает исследование выбранного объекта, который можно назвать основным, и группы вспомогательных объектов. Например, при мониторинге деформаций и осадок в процессе строительства основного сооружения наблюдение ведется за этим основным объектом, но при этом отслеживают окружающие объекты, которые также могут испытывать смещения и осадки.

Совокупность развития ситуаций в этом случае схематически приведена в табл. 2.

В этом случае ситуация А и сценарий 1 соответствуют стационарному геоинформационному моделированию. Ситуации А и сценарии 2—4, Б и сце-

нарии 2, 4, В и сценарии 1—4, Г и сценарии 1—4 требуют проведения геоинформационного ситуационного моделирования. Таким образом, в реальной практике при исследовании совокупности объектов чаще всего требуется проведение ситуационного геоинформационного моделирования. Основой такого моделирования являются пространственные отношения.

На практике пространственные отношения имеют четыре основных вида: иерархические, топологические, геореференциальные [5] и геостатистические [6]. Первые два вида в явной форме задают отношения порядка. Вторые два вида представляют собой неявное знание [7] и требуют достаточно кропотливой обработки для получения явного знания.

Иногда выделяют более мелкие пространственные отношения, такие как тангенциальные [8], конфигурационные, координатные [9].

Иерархические пространственные отношения наиболее ярко проявляются в кадастре. В этой сфере имеет место четкая иерархия кадастровых объектов, которая используется на рынке недвижимости, при управлении землепользованием, в государственном управлении при учете природных и земельных ресурсов. Она не допускает перекрытие объектов одного уровня иерархии.

Самым распространенным типом пространственных отношений при изучении пространственных объектов является иерархический тип, описывающий отношения между элементами, множествами и частями объектов. Это отношения типа "один ко многим".

Однако пространственные отношения включают еще отношения "многие ко многим". Наиболее распространенные пространственные отношения обозначают [10] *ISA* (рис. 1), *AKO*.

Таблица 1

Сценарии развития ситуаций

Среда	Модель объекта	
	Стационарная	Динамическая
Стационарная	Сценарий 1	Сценарий 2
Динамическая	Сценарий 3	Сценарий 4

Таблица 2

Сценарии развития ситуаций при учете совокупности объектов

Ситуация моделирования	Среда стационарная	Среда динамическая
Модель основного объекта стационарная	Сценарий 1	Сценарий 2
Модель основного объекта динамическая	Сценарий 3	Сценарий 4
Совокупность окружающих объектов стационарная	А	Б
Совокупность окружающих объектов динамическая	В	Г

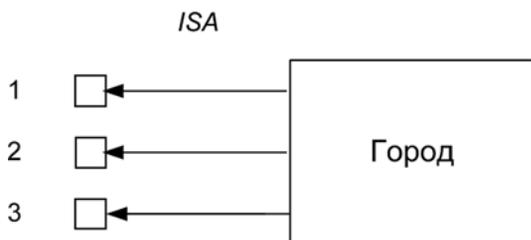


Рис. 1. Пространственные отношения типа ISA:  
1 — сооружение; 2 — переход; 3 — светофор

Отношение классификации *ISA* происходит от английского "is a". Говорят, что множество (класс) классифицирует свои экземпляры (например, "улица есть часть городской территории"). Иногда это отношение именуют "member of". По-русски это может называться "есть" (единственное число) или "суть" (множественное число). Связь *ISA* предполагает, что свойства объекта наследуются от множества. Эти отношения используют при описании городской ситуации.

Обратное отношение — "example of", или "пример". Поэтому процесс порождения элементов из множества называется экземплярцией.

Отношение между множеством и подмножеством *AKO* (рис. 2) происходит от английского "a kind of", например, "Городские районы есть подмножество городской территории".

Отличие *AKO* от отношения *ISA* заключается в том, что *ISA* — отношение "один ко многим", а *AKO* — отношение "многое к многим". С позиций геоинформатики отношение *ISA* применяют для описания ситуаций на картах в крупных масштабах. Отношение *AKO* применяют для описания ситуаций на картах в мелких масштабах.

Применяя пространственные отношения, следует четко различать, какие объекты являются классами, а какие — экземплярами классов. При этом вовсе

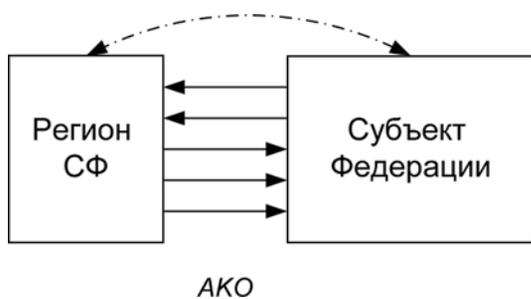


Рис. 2. Пространственные отношения типа AKO



Рис. 3. Отношения холонимии и меронимии

не обязательно, чтобы одно и то же понятие было классом или экземпляром во всех предметных областях. Например, "студент" всегда будет классом в базах знаний типа "студенческая группа" или "вуз", но может быть экземпляром класса учащихся. Наличие отношения классификации еще не говорит о существовании системы классификации, а только служит основой для нее. Исключение составляют те случаи, когда классификация уже создана.

Объект как сложная система состоит из нескольких частей и элементов. Например, город включает улицы, площади, дома, объекты инфраструктуры, инженерные сооружения и т. д. Это определяет еще один тип отношения — "отношение целого и части". Их также называют отношениями меронимии и холонимии (рис. 3).

Отношение меронимии — отношение целого к части ("has part"). Мероним — объект, включающий другой объект как часть, например, "Город включает городские районы. Городская территория включает улицы".

Отношение холонимии — отношение части к целому ("is a part"), например, "Улица часть городской территории". Улица — холоним для городской территории. Городская территория — мероним для улицы.

Говоря об иерархических пространственных отношениях, следует упомянуть об уровнях иерархии и отметить, что существуют семантические отношения на одном уровне иерархии и отношения между уровнями. Отношения на одном уровне соответствуют парадигматическим отношениям, отношения между уровнями соответствуют синтагматическим отношениям. Эти отношения задают семантику ситуационного моделирования в геоинформатике.

Топологические отношения широко используются в картографии и на транспорте. Для описания топологических отношений пространственных объектов широко применяют графовые модели. В этих моделях могут быть использованы разные виды отношений. При этом следует отличать пространственный граф, который содержит пространственную топологию от информационного графа, например семантической сети.

При моделировании топологических отношений используют следующие виды отношений:

- функциональные (определяемые обычно глаголами "производит", "влияет"...);
- количественные (больше, меньше, равно...);
- позиционные (далеко от, близко от, за, под, над...);
- временные (раньше, позже, в течение...);
- атрибутивные (иметь свойство, иметь значение);
- логические (И, ИЛИ, НЕ);
- лингвистические.

Геореферентные отношения [11] представляют собой специфический лингвистически-информационный инструмент фиксации отношений пространственных номенов и понятий, а также геогра-

фических названий. Геореференция в разных формах отражает отношения между пространственно локализованными объектами и информацией об этих объектах. Она часто нелинейна, многозначна и порой бывает неявной. Это означает, что она выражается в категориях неявного знания [7].

В сферу геореференции входят также зависимости между любой документальной информацией (например, карты, космические снимки, биографическая информация) и географической локализацией с помощью местонаименований, кодов места (например, почтовые коды), координат и других методов, описывающих пространственные связи и отношения. Существуют специальные справочники географических названий и так называемые газет-тиры, которые отражают отношение геореференций.

В ГИС геореференция отражается с помощью так называемых ассоциативных связей. Вне ГИС она отражается различными лингвистическими конструкциями.

Геостатистические отношения — отношения, выявляемые методами геостатистики [6, 12]. Развитием и применением методов геостатистики занимаются: экологи, инженеры-нефтяники, гидрологи, почвоведы, геологи, а также статистики. Геостатистика изучает, в первую очередь, пространственные явления и пространственные информационные поля. Геостатистические отношения на основе дискретных измерений в пространстве позволяют создавать непрерывное информационное поле в этом пространстве, например, поле экологической ситуации или поле распределения стоимости недвижимости в пространстве.

На рис. 4 и 5 дается сравнение двух видов интерполяции: пространственной и геостатистической. На рис. 4 дано выражение пространственных отношений на основе пространственной интерполяции. Это пример моделирования ситуации, в которой может быть построена модель объекта исследований. На рисунке приведено плановое положение точек, по которым строится интерполирующая функция поверхности. На основе интерполирующей функции рельефа моделируются изолинии. Изолинии в данном случае — информационная модель ситуации. Интерполяция рельефа, как пример цифрового пространственного моделирования местности, основана на получении регулярной сетки.

Эта сетка получается либо измерением, либо переычислением. На основе регулярной сетки в плановых координатах решается обратная задача для нахождения интерполирующей функции. То есть процесс моделирования ситуации в этом случае требует предобработки исходных пространственных данных.

Горизонталы как изолинии могут проходить через опорные точки, а могут и не проходить. Это обусловлено тем, что функция интерполяции не зависит от расположения исходных точек. Это интер-

поляция по функции. В геостатистике ситуация иная (рис. 5).

Локализация в геостатистике имеет, по крайней мере, два варианта: одно — это просто положение точки в пространстве (которое существует только в абстрактном математическом смысле) и второе — площадь или объем в пространстве вокруг данной точки.

Геостатистическая интерполяция всегда такова, что через точки измерений проходит интерполирующая функция. Это интерполяция по точкам измерений (рис. 5). Она не требует предобработки, как в предыдущем случае, а осуществляет моделирование ситуации и объекта (информационное поле) по исходным данным.

Объектом геостатистического моделирования является информационное поле.

Геостатистические отношения — это отношения между геостатистической локализацией и значением некоего параметра в точке пространства. Геостатистическая локализация отличается от пространственной тем, что кроме пространственных координат измеряется значение некой величины в точке локализации. Для данной точки это может быть концентрация вредных веществ, содержание металла в руде, стоимость недвижимости [13] и т. д. Эти величины не являются точечными, а распределены в пространстве и измеряются как "единица/площадь" или "единица/объем". Единица может быть концентрацией, стоимостью, плотностью и т. д.

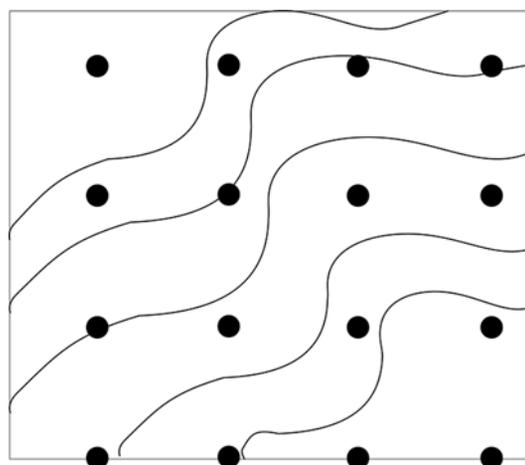


Рис. 4. Пространственная интерполяция рельефа по точкам

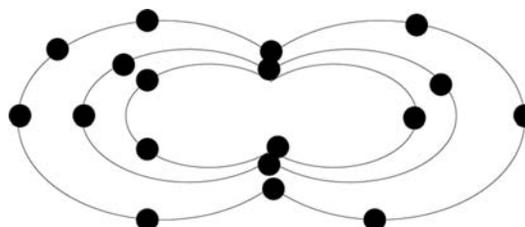


Рис. 5. Геостатистическая интерполяция

Например, значение тематических данных, связанных с областью локализации, может быть средним значением наблюдаемой величины: среднее значение на площадь области, значение на объем области. В последнем случае площадь или объем часто называют "поддержкой" данных. Эта идея тесно связана с идеей поддержки измерений. Геоинформационное моделирование с использованием геостатистики опирается на ряд программных продуктов, наиболее известным из которых является "кригинг". Геостатистика описывает стационарную ситуацию.

### Формализация ситуационного геоинформационного моделирования

Ситуационное моделирование в концептуальном плане соотносится с работами Д. А. Поспелова. В аспекте управления это моделирование соотносится со "школой управления при непредвиденных обстоятельствах". Концепции этой школы, называемой в оригинале *contingency school of management* [14], строятся на том, что для всех случаев жизни нет никакого единственного универсального способа управления. Каждая ситуация уникальна. Не существует единого для всех "лучшего" способа управления во всех ситуациях. Школа управления при непредвиденных обстоятельствах концентрирует внимание на том, что пригодность различных методов управления определяется конкретной ситуацией. Самым эффективным методом управления в конкретной ситуации является метод, кото-

рый более всего соответствует данной ситуации. Задача — найти и суметь реализовать этот метод. Реализацией этого метода в пространственных ситуациях является геоинформационное ситуационное моделирование.

Стационарное геоинформационное моделирование рассматривает объект моделирования как единую систему. Это может быть цифровая модель местности, цифровая модель сооружения, цифровая карта. При визуализации такого моделирования аналогом будет снимок объекта моделирования.

Ситуационное моделирование предполагает моделирование объектов как связанных и зависимо меняющихся систем. При визуализации такого моделирования аналогом будет видеофильм или видеозапись. Структурно ситуационное геоинформационное моделирование опирается на ряд специализированных моделей (рис. 6).

Основой ситуационного моделирования является предварительное создание языка моделирования, т. е. создание различных информационных единиц. Это могут быть графические информационные единицы, единицы информационного взаимодействия, единицы коммуникационного обмена, структурные информационные единицы [15] и информационные семантические единицы [16].

На основе информационных единиц формируются составные информационные модели объекта и ситуации [17]. Понятие ситуации и позиции используется широко [18], например, ситуация и позиция в одном пространстве. Это могут быть: пространственная ситуация и пространственная позиция, экономическая ситуация и экономическая позиция, экологическая ситуация и экологическая позиция.

Возможны комбинации ситуаций и позиций, например, пространственная ситуация и экономическая позиция; пространственная ситуация, экологическая ситуация, пространственная позиция и экономическая позиция. В целом такие комбинации создают основу геоинформационного ситуационного моделирования.

В геоинформатике информационная модель ситуации ISMg имеет вид

$$ISMg = \Phi\{(C_1, C_2, \dots, C_n), (Pt_1, Pt_2, \dots, Pt_m), (A_1, A_2, \dots, A_k)\}. \quad (1)$$

Здесь  $C_i$  — совокупность координатных (пространственных) параметров ( $i = 1, \dots, n$ );  $Pt_i$  — совокупность временных параметров ( $i = 1, \dots, m$ );  $A_i$  — совокупность тематических характеристик ( $i = 1, \dots, k$ ).

При ситуационном моделировании всегда ставится определенная цель. Наличие в модели параметров, необходимых для достижения данной цели  $T_1$ , дает основание такую модель информационной ситуации считать полной по задаче (цели)  $T_1$  и ситуацию моделируемой.

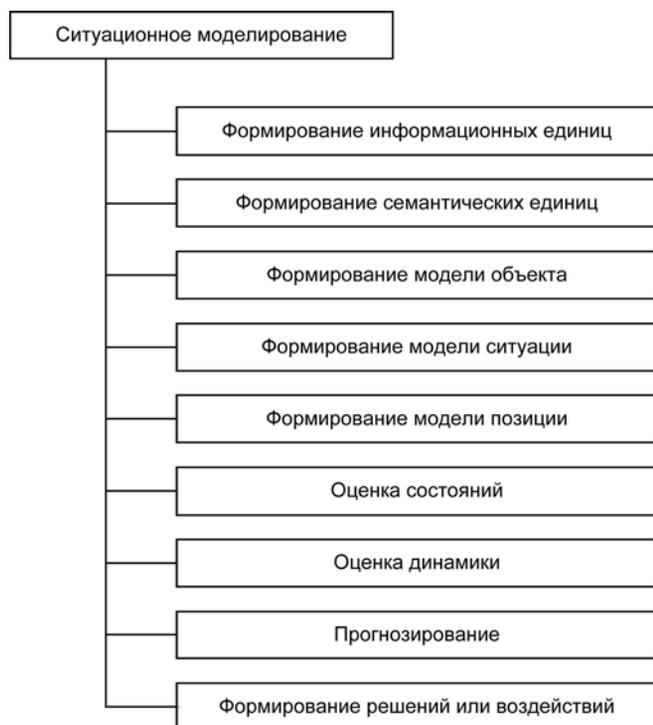


Рис. 6. Этапы ситуационного геоинформационного моделирования

Модель (1) является открытой, и ряд параметров можно дополнять. Для каждого из информационно определяемых параметров "А" должен существовать справочник кодов или классификатор. Это приводит к онтологиям и сближает геоинформационное ситуационное моделирование с технологиями получения знаний.

В целом геоинформационное моделирование в совокупности с информационным подходом и информационным моделированием служит источником получения новых знаний. На основе этих моделей осуществляют прогноз и выработку управленческих решений.

### Выводы

Ситуационное геоинформационное моделирование включает три компонента:

- задание и определение пространственных отношений как основы моделирования и развития событий;
- определение модели микросреды, которая задает ситуацию;
- определение модели объекта с учетом связей и отношений с микросредой и макросредой.

Ситуационное геоинформационное моделирование создает в итоге новые информационные ресурсы и новые знания. Оно позволяет решать широкий круг задач, который с помощью иных методов моделирования решить нельзя. Особое значение ситуационного моделирования в том, что оно используется при поддержке принятия решений, при управлении подвижными объектами. Оно входит в интеллектуальные транспортные системы и интеллектуальные логистические системы.

### Список литературы

1. **Майоров А. А.** Состояние и развитие геоинформатики // *Науки о Земле*. 2012. Вып. 03. С. 11–16.
2. **Максудова Л. Г., Савиных В. П., Цветков В. Я.** Интеграция наук об окружающем мире в геоинформатике // *Исследование Земли из космоса*. 2000. № 1. С. 46–50.
3. **Цветков В. Я.** Геоинформационное моделирование // *Информационные технологии*. 1999. № 3. С. 23–27.
4. **Цветков В. Я.** Пространственные отношения в геоинформатике // *Науки о Земле*. 2012. Вып. 01. С. 59–61.
5. **Hill L. L.** *Georeferencing: The Geographic Associations of Information*. London: MIT Press Cambridge, Massachusetts, 2009. 272 p.
6. **Цветков В. Я.** Геостатистика // *Геодезия и аэрофотосъемка*. 2007. № 3. С. 174–184.
7. **Kimble C.** Knowledge management, codification and tacit knowledge // *Information Research*, 2013. N 18 (2). P. 577.
8. **Кулагин В. П., Цветков В. Я.** Геознание: представление и лингвистические аспекты // *Информационные технологии*. 2013. № 12. С. 2–9.
9. **Цветков В. Я.** Виды пространственных отношений // *Успехи современного естествознания*. 2013. № 5 С. 138–140.
10. **Цветков В. Я.** *Общая геоинформатика*. Saarbrücken: LAP LAMBERT Academic Publishing GmbH & Co. KG. 2013. 288 p.
11. **Moritz T.** Geo-referencing the natural and cultural world, past and present: Towards building a distributed, peer-reviewed gazetteer system // *Digital Gazetteer Information Exchange Workshop*. Smithsonian Institute, October 12–14, 1999.
12. **Pichler G.** *Computer-Programme der Geostatistik*. Master's Thesis. Institut für Statistik, Technische Universität, Graz, Austria, 1982.
13. **Майоров А. А., Матерухин А. В.** Геоинформационный подход к задаче разработки инструментальных средств массовой оценки недвижимости // *Геодезия и аэрофотосъемка*. 2011. № 4. С. 92–97.
14. **Encyclopedia of Management**. URL: <http://www.enotes.com/management-encyclopedia/management-thought>
15. **Tsvetkov V. Ya.** Information objects and information Units // *European Journal of Natural History*. 2009. N 2. P. 99.
16. **Tsvetkov V. Ya.** Semantic Information Units as L. Floridi's Ideas Development // *European Researcher*. 2012. Vol. (25), N 7. P. 1036–1041.
17. **Tsvetkov V. Ya.** Information Situation and Information Position as a Management Tool // *European Researcher*. 2012. Vol. (36), N 12-1. P. 2166–2170.
18. **Соловьев И. В.** Применение модели информационной ситуации в геоинформатике // *Науки о Земле*. 2012. № 01. С. 54–58.

УДК 004.942

**А. И. Юрин**, канд. техн. наук, доц., **М. И. Красивская**, ст. преподаватель,  
**А. В. Дмитриев**, канд. техн. наук, доц., **Г. Ю. Злодеев**, аспирант,  
Московский институт электроники и математики Национального исследовательского университета  
Высшая школа экономики, e-mail: ayurin@hse.ru

## Применение виртуальных лабораторных стендов в образовательном процессе

*Рассмотрены особенности применения виртуальных лабораторных стендов в образовательном процессе. Приведен пример реализации виртуального лабораторного стенда в среде NI LabVIEW.*

**Ключевые слова:** лабораторный стенд, NI LabVIEW

A. I. Yurin, M. I. Krasivskaya, A. V. Dmitriev, G. Yu. Zlodeev

## Virtual Laboratory Stands Application in Education

*The features of virtual laboratory stands in education. An example of the virtual laboratory stand developed with NI LabVIEW.*

**Keywords:** laboratory stand, NI LabVIEW

На сегодняшний день многие учебные заведения используют для подготовки бакалавров, специалистов и магистров новые формы обучения с использованием компьютерных технологий, базирующихся на современных программных продуктах. Организация учебной деятельности на основе образовательных стандартов требует, чтобы теоретические выводы студенты делали при анализе результатов собственной практической работы [1]. Лабораторные работы являются основой для обобщения и закрепления теоретических знаний и получения необходимых профессиональных навыков. Поэтому многие образовательные программы предусматривают в своем составе лабораторные практикумы.

При реализации некоторых актуальных направлений подготовки, например связанных с наукоемким производством, могут потребоваться практикумы, направленные на приобретение навыков работы со сложным, дорогостоящим, уникальным оборудованием. При этом непосредственное использование такого оборудования в учебном процессе может быть невозможным или экономически неоправданным. Кроме того, возникает задача организации практикумов для студентов различных форм обучения (очная, очно-заочная, заочная).

В настоящее время большинство образовательных учреждений оснащены современной компьютерной техникой, поэтому одним из актуальных направлений решения обозначенной выше задачи является реализация лабораторных практикумов с использованием специализированных виртуальных лабораторных стендов.

Виртуальный лабораторный стенд представляет собой программное обеспечение, позволяющее имитировать реальные процессы и максимально близко отражать принципы, режимы и порядок работы соответствующего оборудования. Реализация такого программного обеспечения позволяет проводить лабораторные работы без использования реального оборудования.

Однако выполнение лабораторных работ с использованием виртуальных стендов не должно отменять необходимости освоения студентами реального научно-исследовательского оборудования. Такие лабораторные работы могут быть использованы как демонстрационные, тренировочные или контрольные перед реальной лабораторной работой. Это целесообразно, например, если лабораторная работа предполагает использование сложного оборудования, и неправильные действия студента потенциально могут привести к его неисправности.

Большим достоинством применения виртуальных лабораторных стендов является возможность их использования для дистанционного обучения студентов при отсутствии возможности работы в реальной лаборатории. Как правило, единственное, что необходимо для проведения дистанционной лабораторной работы, это наличие персонального компьютера с доступом в сеть Internet и комплекта методических указаний.

Современный уровень развития информационных технологий и множество существующих языков программирования и средств разработки приложений создают большой выбор путей построения виртуальных лабораторных стендов [2]. Помимо универсальных средств, существует ряд специализированных сред разработки, ориентированных на создание ПО для работы с контрольно-измерительным оборудованием и создания измерительных систем. Наиболее широкое распространение в этой области получила среда графического программирования National Instruments (NI) LabVIEW (США). В этой среде реализуется так называемая технология виртуальных приборов (ТВП), при использовании которой конечные функции разрабатываемой системы во многом определяются программным обеспечением. Этот подход обеспечивает значительную гибкость решений и дает возможность на одной аппаратной базе реализовать различные средства измерений, в зависимости от конкретных задач. При этом NI LabVIEW позволяет реализовать графический интерфейс прикладного программного обеспечения, имитирующий внешний вид классического измерительного прибора.

ТВП предоставляет также широкие возможности для создания прототипов и моделей измерительных устройств, в том числе с использованием записанных массивов реальной измерительной информации. Это обуславливает широкое применение ТВП не только для решения задач автоматизации измерений, но и для создания виртуальных лабораторных стендов, основанных на использовании моделей реального измерительного оборудования [3].

На данный момент имеется острая необходимость разработки виртуальных лабораторных стендов для инженерных образовательных программ. Это обусловлено тем, что учебные заведения удалены от предприятий, не всегда имеют необходимое оборудование, а закупать оборудование только для учебных целей зачастую нецелесообразно. В связи с этим применение виртуальных лабораторных стендов является наиболее удобным и экономически выгодным для подготовки специалистов наукоемких отраслей, в частности для nanoиндустрии.

В качестве примера программной реализации виртуального лабораторного стенда для специалистов в области нанотехнологий можно привести "Виртуальный конфокальный микроскоп" (ВКМ), разработанный сотрудниками Московского инсти-

тута электроники и математики Высшей школы экономики совместно со Всероссийским научно-исследовательским институтом оптико-физических измерений по заказу Метрологического центра РосНАНО. Программное обеспечение ВКМ (см. рисунок на четвертой стороне обложки) разработано в среде NI LabVIEW, предназначено для образовательных целей и эмулирует работу с конфокальным сканирующим микроскопом VCM 200A (Veeco Instruments Inc). Назначением Veeco VCM 200A является измерение параметров профиля поверхности и шероховатости отражающих объектов. При создании модели были использованы видеоролики с записью работы микроскопа, а также изображения и реальные профили объектов, полученные при работе с конфокальным микроскопом в различных режимах. Кроме того, в графический интерфейс пользователя программного обеспечения (ПО) ВКМ включен ряд основных элементов управления и индикации, аналогичных присутствующим в ПО реального конфокального микроскопа.

Программное обеспечение ВКМ позволяет эмулировать основные операции работы с конфокальным микроскопом.

При выборе нужного микрообъектива его смена сопровождается соответствующим аудиовизуальным рядом. В соответствии с выбранным увеличением выводится изображение исследуемого объекта в двухмерном или трехмерном представлении. С помощью ручки Z в двухмерном режиме осуществляется регулировка вертикальной позиции для получения сфокусированного изображения. При правильной фокусировке, после нажатия кнопки "Сканирование", выводится конфокальное изображение объекта. На конфокальном изображении объекта может быть проведена линия сечения и получен профиль. При этом программное обеспечение контролирует корректность проведения линии сечения. После получения профиля он может быть использован для выполнения измерений. В процессе выполнения измерений их результаты можно записывать в специальную предварительную таблицу.

В трехмерном режиме с помощью компьютерной мыши можно осуществлять поворот представления объекта.

В ПО также реализован режим "Обработка". В этом режиме возможно формирование серий записанных измерений, редактирование данных, расчет основных параметров полученных серий измерений, формирование сводных данных. Полученные результаты измерений и расчетов могут быть записаны в файл протокола. В протокол также заносят дату и место проведения работ, фамилию оператора. Протокол формируется в виде HTML-документа для последующей работы с помощью Интернет-браузера или офисных приложений.

Для оценки точности результатов измерений был разработан и реализован программными сред-

ствами алгоритм расчета суммарной неопределенности измерений [4]. Стандартная неопределенность типа  $A$  ( $u_A$ ) выражается путем статистического анализа результатов многократных измерений:

$$u_A = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2},$$

где  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ .

Для оценки стандартной неопределенности типа  $B$  ( $u_B$ ) используются значения погрешности применяемых при поверке и калибровке мер ( $\Delta_M$ ):

$$u_B = \frac{\Delta_M}{\sqrt{3}}.$$

Суммарную неопределенность  $u_C$  вычисляют по формуле

$$u_C = \sqrt{u_A^2 + u_B^2}.$$

Разработанное программное обеспечение ВКМ можно использовать также в качестве виртуального тренажера при подготовке специалистов в области нанометрологии. При этом исключается использование реального оборудования — как моделируемого микроскопа, так и эталонного и вспомогательного оборудования, применяемого при проведении процедур поверки или калибровки, которое имеет высокую стоимость и не может быть использовано как учебное.

Таким образом, разработанное программное обеспечение ВКМ позволяет будущим специалистам получать необходимые навыки работы с конфокальным микроскопом, определять метрологические характеристики без использования реального прибора или расходования ресурса.

#### Список литературы

1. [www.fgosvo.ru](http://www.fgosvo.ru)
2. Стародубцев В. А. Использование современных компьютерных технологий в инженерном образовании: учеб. пособие. Томск: Изд-во Томского политехнического университета, 2008.
3. [www.labview.ru](http://www.labview.ru)
4. РМГ 43—2001. ГСИ. Применение "Руководства по выражению неопределенности измерений".

---

---

#### Адрес редакции:

107076, Москва, Стромынский пер., 4

Телефон редакции журнала (499) 269-5510

E-mail: [it@novtex.ru](mailto:it@novtex.ru)

Дизайнер Т.Н. Погорелова. Технический редактор Е.В. Конова.

Корректор Е.В. Комиссарова.

Сдано в набор 06.04.2014. Подписано в печать 21.05.2014. Формат 60×88 1/8. Бумага офсетная.

Усл. печ. л. 8,86. Заказ IT614. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Оригинал-макет ООО "Авансед солюшнз". Отпечатано в ООО "Авансед солюшнз".

119071, г. Москва, Ленинский пр-т, д. 19, стр. 1.

---