

ИНФОРМАЦИОННО- УПРАВЛЯЮЩИЕ СИСТЕМЫ

НАУЧНЫЙ ЖУРНАЛ

2(81)/2016

2(81)/2016

INFORMATSIONNO- UPRAVLIAIUSHCHIE SISTEMY (INFORMATION AND CONTROL SYSTEMS)

REFEREED EDITION

Founder
«Information and Control Systems», Ltd.

Publisher
Saint-Petersburg State University
of Aerospace Instrumentation

Editor-in-Chief
M. Sergeev
Dr. Sc., Tech., Professor, St. Petersburg, Russia

Deputy Editor-in-Chief
E. Krouk
Dr. Sc., Tech., Professor, St. Petersburg, Russia

Executive secretary
O. Muravtsova

Editorial Council
C. Christodoulou
PhD, Professor, Albuquerque, New Mexico, USA
L. Chubraeva
RAS Corr. Member, Dr. Sc., Tech., Professor, St. Petersburg, Russia
L. Fortuna
PhD, Professor, Catania, Italy
A. Fradkov
Dr. Sc., Tech., Professor, St. Petersburg, Russia
V. Kozlov
Dr. Sc., Tech., Professor, St. Petersburg, Russia
B. Meyer
Dr. Sc., Professor, Zurich, Switzerland
A. Ovodenko
Dr. Sc., Tech., Professor, St. Petersburg, Russia
Y. Podoplyokin
Dr. Sc., Tech., Professor, St. Petersburg, Russia
Yu. Shokin
RAS Academician, Dr. Sc., Phys.-Math., Novosibirsk, Russia
V. Simakov
Dr. Sc., Tech., Professor, Moscow, Russia
V. Vasilev
RAS Corr. Member, Dr. Sc., Tech., Professor, St. Petersburg, Russia
R. Yusupov
RAS Corr. Member, Dr. Sc., Tech., Professor, St. Petersburg, Russia

Editorial Board
V. Anisimov
Dr. Sc., Tech., Professor, St. Petersburg, Russia
B. Bezruchko
Dr. Sc., Phys.-Math., Saratov, Russia
N. Blaunstein
Dr. Sc., Phys.-Math., Professor, Beer-Sheva, Israel
A. Dudin
Dr. Sc., Tech., Professor, Minsk, Belarus
I. Dumer
PhD., Professor, Riverside, USA
V. Khimenko
Dr. Sc., Tech., Professor, St. Petersburg, Russia
G. Maltsev
Dr. Sc., Tech., Professor, St. Petersburg, Russia
V. Melekhin
Dr. Sc., Tech., Professor, St. Petersburg, Russia
A. Shalyto
Dr. Sc., Tech., Professor, St. Petersburg, Russia
A. Shepeta
Dr. Sc., Tech., Professor, St. Petersburg, Russia
A. Smirnov
Dr. Sc., Tech., Professor, St. Petersburg, Russia
Z. Yuldashev
Dr. Sc., Tech., Professor, St. Petersburg, Russia
A. Zeifman
Dr. Sc., Phys.-Math., Vologda, Russia

Editor: A. Larionova
Proofreader: T. Zvertanovskaia
Design: N. Karavaeva, M. Chernenko
Layout and composition: N. Karavaeva, J. Umnitsina

Contact information
The Editorial and Publishing Center, SUAI
67, B. Morskaia, 190000, St. Petersburg, Russia
Website: <http://i-us.ru/en>, E-mail: ius.spb@gmail.com
Tel.: +7 - 812 494 70 02

THEORETICAL AND APPLIED MATHEMATICS

Kuchmin A. Yu. *A Nonlinear Programming Method with Arbitrary Restrictions* 2

INFORMATION PROCESSING AND CONTROL

Proskurin A. V., Favorskaya M. N. *Automatic Image Annotation Based on Homogeneous Textual-Visual Groups* 11

Krasilnikov N. N., Krasilnikova O. I. *Study of Critical Values of On-Screen Parallax* 19

INFORMATION AND CONTROL SYSTEMS

Egorov I. V., Melekhin V. F. *Methods and Tools for Structural Block Reliability Analysis with Reservation and Periodic Information Recovery at Various Stages of Computing System Design* 26

Burakov M. V., Konovalov A. S. *Fuzzy Control for Anti-Lock Automobile Brake Systems* 35

INFORMATION CODING AND TRANSMISSION

Kul'minskii D. D., Ponomarenko V. I., Prokhorov M. D., Bezruchko B. P. *Communication System based on Generalized Chaotic Synchronization* 42

Kutuzov O. I., Tatarnikova T. M. *Evaluation and Comparison of Queues in Classical and Fractal Queuing Systems* 48

Trofimov A. N., Taubin F. A. *Information Theory Analysis of Multilevel Flash Memory. Part 1. Channel Model and Random Coding Bounds* 56

Basov O. O., Kipyatkova I. S., Saveliev A. I., Saitov I. A. *Polymodal Information Encoding Models* 68

INFORMATION CHANNELS AND MEDIUM

Maltsev G. N., Kozinov I. A. *Transfer of Hyperspectral Video Data of Earth Remote Sensing via Radio Channels of Limited Capacity* 74

INFORMATION AND MEASURING SYSTEMS

Kulakov S. V., Balysheva O. L., Kludzin V. V. *Methods of Acousto-Optics in Studying the Properties of Materials for Acousto-Optic and Acousto-Electronic Devices* 84

Kaplin A. Y., Stepanov M. G. *Analysis of Algorithm for Complex Processing of Goniometric Information in a Moving Object Navigation System* 89

INFORMATION INSTRUMENTATION AND EDUCATION

Kulikov G. G., Antonov V. V., Shilina M. A., Fakhruullina A. R. *Structuring Domain Content for Further Data Mining: an example of Forming Structured Content for Education-and-Production Activity* 95

BRIEF SCIENTIFIC REPORTS

Balonin N. A., Sergeev M. B. *Propus Matrices 92 and 116* 101

Ziatdinov S. I. *Analysis of Linear Systems on the base of Transient Characteristics* 104

INFORMATION ABOUT THE AUTHORS

107

2(81)/2016

РЕЦЕНЗИРУЕМОЕ ИЗДАНИЕ

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

Учредитель
ООО «Информационно-управляющие системы»

Издатель
Санкт-Петербургский государственный университет
аэрокосмического приборостроения

Главный редактор
М. Б. Сергеев,
д-р техн. наук, проф., С.-Петербург, РФ

Зам. главного редактора
Е. А. Крук,
д-р техн. наук, проф., С.-Петербург, РФ

Ответственный секретарь
О. В. Муравцова

Редакционный совет:

Председатель А. А. Оводенко,
д-р техн. наук, проф., С.-Петербург, РФ
В. Н. Васильев,
чл.-корр. РАН, д-р техн. наук, проф., С.-Петербург, РФ
В. Н. Козлов,
д-р техн. наук, проф., С.-Петербург, РФ
К. Кристофолу,
д-р наук, проф., Альбукерке, Нью-Мексико, США
Б. Мейер,
д-р наук, проф., Цюрих, Швейцария
Ю. Ф. Подоплёкин,
д-р техн. наук, проф., С.-Петербург, РФ
В. В. Симаков,
д-р техн. наук, проф., Москва, РФ
Л. Фортуну,
д-р наук, проф., Катания, Италия
А. Л. Фрадков,
д-р техн. наук, проф., С.-Петербург, РФ
Л. И. Чубраева,
чл.-корр. РАН, д-р техн. наук, С.-Петербург, РФ
Ю. И. Шокин,
акад. РАН, д-р физ.-мат. наук, проф., Новосибирск, РФ
Р. М. Юсупов,
чл.-корр. РАН, д-р техн. наук, проф., С.-Петербург, РФ

Редакционная коллегия:

В. Г. Анисимов,
д-р техн. наук, проф., С.-Петербург, РФ
Б. П. Безручко,
д-р физ.-мат. наук, проф., Саратов, РФ
Н. Блаунштейн,
д-р физ.-мат. наук, проф., Беэр-Шева, Израиль
А. Н. Дудин,
д-р физ.-мат. наук, проф., Минск, Беларусь
И. И. Думер,
д-р наук, проф., Риверсайд, США
А. И. Зейфман,
д-р физ.-мат. наук, проф., Вологда, РФ
Г. Н. Мальцев,
д-р техн. наук, проф., С.-Петербург, РФ
В. Ф. Мелехин,
д-р техн. наук, проф., С.-Петербург, РФ
А. В. Смирнов,
д-р техн. наук, проф., С.-Петербург, РФ
В. И. Хименко,
д-р техн. наук, проф., С.-Петербург, РФ
А. А. Шалыто,
д-р техн. наук, проф., С.-Петербург, РФ
А. П. Шепета,
д-р техн. наук, проф., С.-Петербург, РФ
З. М. Юлдашев,
д-р техн. наук, проф., С.-Петербург, РФ

Редактор: А. Г. Ларионова
Корректор: Т. В. Звертановская
Дизайн: Н. Н. Караваева, М. Л. Черненко
Компьютерная верстка: Н. Н. Караваева, Ю. В. Умницына

Адрес редакции: 190000, Санкт-Петербург,
Б. Морская ул., д. 67, ГУАП, РИЦ
Тел.: (812) 494-70-02, e-mail: ius.spb@gmail.com, сайт: http://i-us.ru

Журнал зарегистрирован в Министерстве РФ по делам печати,
телерадиовещания и средств массовых коммуникаций.
Свидетельство о регистрации ПИ № 77-12412 от 19 апреля 2002 г.
Перерегистрирован в Роскомнадзоре.
Свидетельство о регистрации ПИ № ФС77-49181 от 30 марта 2012 г.

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий,
в которых должны быть опубликованы основные научные результаты диссертации
на соискание ученой степени доктора и кандидата наук».

© Коллектив авторов, 2016

ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ МАТЕМАТИКА

Кучмин А. Ю. Об одном методе нелинейного программирования
с произвольными ограничениями

2

ОБРАБОТКА ИНФОРМАЦИИ И УПРАВЛЕНИЕ

Проскурин А. В., Фаворская М. Н. Автоматическое аннотирование
изображений на основе однородных текстово-визуальных групп

11

Красильников Н. Н., Красильникова О. И. Исследование критического
значения экранного параллакса

19

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

Егоров И. В., Мелехин В. Ф. Методы и средства анализа
надежности структурных блоков с резервированием и периодическим
восстановлением информации на различных этапах проектирования
вычислительных систем

26

Бураков М. В., Коновалов А. С. Нечеткое управление автомобильной
антиблокировочной системой

35

КОДИРОВАНИЕ И ПЕРЕДАЧА ИНФОРМАЦИИ

Кульминский Д. Д., Пономаренко В. И., Прохоров М. Д.,
Безручко Б. П. Система передачи информации, основанная
на обобщенной хаотической синхронизации

42

Кутузов О. И., Татарникова Т. М. К оценке и сопоставлению
очереди классических и фрактальных систем массового обслуживания

48

Трофимов А. Н., Таубин Ф. А. Теоретико-информационный
анализ многоуровневой flash-памяти. Часть 1. Модель канала
и границы случайного кодирования

56

Басов О. О., Кипяткова И. С., Савельев А. И., Саитов И. А. Модели
кодирования полимодальной информации

68

ИНФОРМАЦИОННЫЕ КАНАЛЫ И СРЕДЫ

Мальцев Г. Н., Козинков И. А. Передача гиперспектральных
видеоданных дистанционного зондирования Земли по радиоканалам
с ограниченной пропускной способностью

74

ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ СИСТЕМЫ

Кулаков С. В., Балышева О. Л., Кудзин В. В. Методы акустооптики
в исследовании свойств материалов для акустооптических
и акустоэлектронных устройств

84

Каплин А. Ю., Степанов М. Г. Анализ алгоритма комплексной
обработки угломерной информации в навигационной системе
подвижного объекта

89

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ОБРАЗОВАНИЕ

Куликов Г. Г., Антонов В. В., Шилина М. А., Фахруллина А. Р.
Структурирование контента предметной области для дальнейшего
интеллектуального анализа. Пример формирования структури-
рованного контента учебно-производственной деятельности

95

КРАТКИЕ СООБЩЕНИЯ

Балонин Н. А., Сергеев М. Б. Матрицы Пропус 92 и 116

101

Зиятдинов С. И. Анализ линейных систем на основе переходных
характеристик

104

СВЕДЕНИЯ ОБ АВТОРАХ

107

Сдано в набор 04.03.16. Подписано в печать 21.04.16. Формат 60×84/8.
Бумага офсетная. Гарнитура SchoolBookC. Печать офсетная.
Усл. печ. л. 13,0. Уч.-изд. л. 16,3. Тираж 1000 экз (1-й завод 150 экз). Заказ 188.
Оригинал-макет изготовлен в редакционно-издательском центре ГУАП.
190000, Санкт-Петербург, Б. Морская ул., 67.
Отпечатано с готовых диапозитивов в редакционно-издательском центре ГУАП.
190000, Санкт-Петербург, Б. Морская ул., 67.

Журнал распространяется по подписке. Подписку можно оформить
через редакцию, а также в любом отделении связи по каталогу «Роспечать»:
№ 48060 — годовой индекс, № 15385 — полугодовой индекс.

ОБ ОДНОМ МЕТОДЕ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ С ПРОИЗВОЛЬНЫМИ ОГРАНИЧЕНИЯМИ

А. Ю. Кучмин^а, канд. техн. наук, ведущий научный сотрудник

^аИнститут проблем машиноведения РАН, Санкт-Петербург, РФ

Введение: при решении современных задач в областях моделирования и идентификации нелинейных и нестационарных систем, а также при анализе и фильтрации данных экспериментов все большее распространение получают методы оптимизации и, в частности, нелинейного программирования. Эти задачи относятся к классу задач многомерной оптимизации при ограничениях. К сожалению, в настоящий момент не существует эффективных численных методов для решения подобных задач в случае, когда число ограничений m многократно превосходит число варьируемых параметров n . **Цель:** разработать новые методы и алгоритмы для решения задач многопараметрической оптимизации при ограничениях. **Результаты:** разработан новый метод нелинейного программирования, который назван «метод многомерных оболочек». Этот метод позволяет решать задачи условной многопараметрической оптимизации для случая $n \ll m$. Он использует как необходимые, так и достаточные условия оптимальности для целевых функций, имеющих вид метрик. Показано, что если единственный минимум (максимум) метрики находится вне области допустимых значений, образованной ограничениями, тогда решение задачи будет находиться на границе области допустимых решений. Тогда задача может иметь несколько решений, которые будут соответствовать одинаковому минимальному (максимальному) значениям целевой функции, и все эти решения могут быть получены с помощью метода многомерных оболочек. Основой данного метода является построение области допустимых решений путем анализа ограничений произвольного вида и их аппроксимации кусочно-полиномиальными ограничениями или рациональными дробями, что позволяет свести исходную задачу к набору взаимосвязанных подзадач с линейными ограничениями, формирующими выпуклые многомерные оболочки, что существенно сужает область поиска. **Практическая значимость:** предложенный метод нелинейного программирования апробирован и найдет применение при решении задач условной оптимизации в областях физической метрологии, оптимальных систем управления и логико-интервальных систем. Данный алгоритм может быть улучшен путем разработки более эффективных алгоритмов для новых типов ограничений.

Ключевые слова — нелинейное программирование, многопараметрическая оптимизация.

Введение

Многие задачи синтеза законов управления, моделирования сложных систем и идентификации объектов могут быть сформулированы как задачи условной оптимизации с ограничениями различных типов. Часто в качестве ограничений выступают уравнения, описывающие исследуемый объект. Они имеют кусочно-полиномиальный, кусочно-экспоненциальный, интервальный вид или являются рациональными дробями.

Такой подход имеет ряд преимуществ при более тонком анализе особенностей различных объектов. Например, в задачах фильтрации экспериментальных данных он позволяет более эффективно разделять физические свойства испытываемого образца и влияние сторонних факторов стенда и внешней среды. А для задач управления и идентификации сочетание оптимизационного подхода и моделей, заданных в виде систем неравенств, может стать аксиоматической альтернативой нечетким системам и нейронным сетям.

Развитие этого направления сдерживает отсутствие эффективных численных методов для решения задач условной оптимизации для большого количества варьируемых переменных n , когда число ограничений m многократно превосходит n .

Цель данной статьи — рассказать о новых идеях в области методов для решения задач условной оптимизации без перехода к задаче безусловной оптимизации, в частности, о разработанном новом методе многомерных оболочек (Multidimensional Patchwork Shell Method).

Постановка задачи условной оптимизации

В математической постановке задача минимизации (максимизации) формулируется в следующем виде:

$$\begin{aligned} J &= \min / \max / f(\mathbf{x}) \text{ при} \\ \mathbf{A}_1 \mathbf{x} &= \mathbf{B}_1, \mathbf{A}_2 \mathbf{x} \leq \mathbf{B}_2, \mathbf{C}_1(\mathbf{x}) = 0, \\ \mathbf{C}_2(\mathbf{x}) &\leq 0, \mathbf{L}_1(\mathbf{x}) = 0, \mathbf{L}_2(\mathbf{x}) \leq 0, \end{aligned} \quad (1)$$

где J — критерий; $\mathbf{x}$ — вектор варьируемых переменных; f — целевая функция; $\mathbf{A}_1, \mathbf{B}_1, \mathbf{A}_2, \mathbf{B}_2$ — матрицы линейных ограничений, $\mathbf{C}_1, \mathbf{C}_2$ — гладкие матричные функции от соответствующих аргументов; $\mathbf{L}_1, \mathbf{L}_2$ — матричные функции с разрывами первого и второго рода от соответствующих аргументов.

Классический подход в решении задач условной оптимизации — это сведение ее к эквивалентной задаче безусловной оптимизации методом множителей Лагранжа [1] с дальнейшим применением разработанного инструментария

методов, например градиентного [2] или случайного [3] поисков.

Недостатки этого подхода заключаются в следующем.

1. Применение метода множителей Лагранжа корректно, если доказано существование седловой точки для модифицированной целевой функции.

2. Не доказана корректность метода множителей Лагранжа для общего случая нелинейных ограничений типа неравенство.

3. Общее число p варьируемых переменных модифицированной целевой функции многократно увеличивается: $p = n + m_p + 2m_n$, где n — число переменных в задаче условной оптимизации; m_p — количество ограничений типа равенство; m_n — количество ограничений типа неравенство. При этом область допустимых решений изменяется от $-\infty$ до $+\infty$ по всем переменным.

4. Большое количество ограничений разного типа может привести к появлению многих экстремумов, даже для случая монотонной исходной целевой функции f .

5. Ограничения с разрывами не позволяют использовать градиентные методы поиска.

Данные недостатки могут быть продемонстрированы на простом примере: необходимо найти минимум целевой функции

$$J = \min_x \{f\}, f = ax^2 + bx + c \text{ при } ax^2 + bx + c \geq d, \quad (2)$$

где J — функционал; f — целевая функция; x — варьируемая переменная; a, b, c, d — постоянные коэффициенты.

Решим эту задачу, анализируя ограничения и целевую функцию. Найдем интервалы аргумента, которые соответствуют ограничению в (2)

$$ax^2 + bx + c - d = 0;$$

$$x_{1,2} = \frac{-b \mp \sqrt{b^2 - 4a(c-d)}}{2a};$$

$$-\infty \leq x \leq +\infty, \text{ если } b^2 - 4a(c-d) < 0 \text{ и } a > 0;$$

$$x \in \emptyset, \text{ если } b^2 - 4a(c-d) < 0 \text{ и } a < 0;$$

$$x \leq x_1, x \geq x_2, \text{ если } b^2 - 4a(c-d) \geq 0 \text{ и } a > 0;$$

$$x_1 \leq x \leq x_2, \text{ если } b^2 - 4a(c-d) \geq 0 \text{ и } a < 0, \quad (3)$$

где $x_{1,2}$ — корни уравнения.

Исследуем целевую функцию на монотонность для всех интервалов аргумента, полученных в (3).

1. Для случая $b^2 - 4a(c-d) < 0$ и $a > 0$ целевая функция f монотонно убывает на интервале $x < -b/2a$, на интервале $x \geq -b/2a$ функция монотонно возрастает. Отсюда следует, что f имеет глобальный минимум в точке $x = -b/2a$.

2. Для случая $b^2 - 4a(c-d) < 0$ и $a < 0$ аргумент принадлежит пустому множеству, и задача не имеет решения.

3. Для случая $a > 0$ и $b^2 - 4a(c-d) \geq 0$ целевая функция f монотонно убывает на интервале $x \leq x_1$, на интервале $x \geq x_2$ функция монотонно возрастает. Отсюда следует, что f имеет два минимума с одинаковым значением f в точках x_1 и x_2 .

4. Для случая $a < 0$ и $b^2 - 4a(c-d) \geq 0$ целевая функция f монотонно возрастает на интервале $x_1 \leq x < -b/2a$, на интервале $-b/2a \leq x \leq x_2$ функция монотонно убывает. Отсюда следует, что f имеет два минимума с одинаковым значением в точках x_1 и x_2 .

Количество варьируемых переменных осталось неизменным, а область поиска была существенно сокращена для случаев 2, 3 и 4. В случае 1 было доказано, что ограничения можно не учитывать.

Для решения задачи (2) воспользуемся переходом от условной оптимизации к безусловной с применением метода множителей Лагранжа:

$$\Phi = \min_{x, \lambda, \xi} \{g\},$$

$$g(x, \lambda, \xi) = f + \lambda(-ax^2 - bx - c + d + \xi^2), \quad (4)$$

где Φ — критерий безусловной оптимизации; λ — множитель Лагранжа; ξ — свободный параметр для учета неравенства.

Если доказано существование седловой точки для g , то справедлива следующая система уравнений:

$$\begin{cases} \frac{\partial g}{\partial x} = 0 \\ \frac{\partial g}{\partial \lambda} = 0 \\ \frac{\partial g}{\partial \xi} = 0 \end{cases} \Rightarrow \begin{cases} 2ax + b - \lambda 2ax - \lambda b = 0 \\ -ax^2 - bx - c + d + \xi^2 = 0 \\ 2\lambda \xi = 0 \end{cases} \Rightarrow \begin{cases} \lambda = 1 \\ x_{1,2} = \frac{-b \mp \sqrt{b^2 - 4a(c-d)}}{2a} \\ \xi = 0 \end{cases} \quad (5)$$

Полученное решение (5) не описывает случай 1, когда $b^2 - 4a(c-d) < 0$ и $a > 0$. Численное решение данной задачи градиентными методами даст только одно значение аргумента, а в случае $b^2 - 4a(c-d) < 0$ и $a < 0$ правильный ответ может быть получен только аналитически.

Количество варьируемых переменных увеличилось в 3 раза, область поиска равна R^3 .

Для использования метода множителей Лагранжа необходимо доказать существование седловой точки модифицированной целевой функции, что в случае большого количества переменных является сложной задачей, сравнимой по трудоемкости с нахождением глобального оптимума.

Данный пример показывает преимущества методов, основанных на анализе ограничений, по сравнению с методами, использующими переход от условной оптимизации к безусловной, а также демонстрирует нетождественность такого перехода.

Уменьшение числа переменных путем анализа линейных ограничений типа равенство

Ограничения типа равенство можно использовать для редукции числа варьируемых переменных и исключения ограничений типа равенство, которые дублируют друг друга.

Рассмотрим ограничения

$$\mathbf{A}_1 \mathbf{x} = \mathbf{B}_1, \quad (6)$$

где матрица $\mathbf{A}_1$ имеет размерности $m_p \times n$; вектор $\mathbf{x}$ имеет размерность $n \times 1$, а $\mathbf{B}_1$ имеет размерности $m_p \times 1$. Здесь возможны три случая.

1. Если число линейно независимых уравнений больше n , то ограничения противоречивы и задача не имеет решения.

2. Если число линейно независимых уравнений равно n и найденное значение $\mathbf{x}_0 = \mathbf{B}_1 \mathbf{A}_1^{-1}$ не противоречит другим ограничениям, то $\mathbf{x}_0$ — это единственное решение задачи. Если $\mathbf{x}_0$ противоречит другим ограничениям, то задача не имеет решения.

3. Если число линейно независимых уравнений n_1 меньше n , тогда необходимо определить базисные ограничения, а все линейно зависимые ограничения должны быть исключены. Полученная система уравнений позволяет определить линейную замену варьируемых переменных и сократить их число до $n - n_1$.

Для случая 3 замена переменных может быть выполнена по рекуррентной формуле

$$z_{n-n_1+k} = b_k - \sum_{i=1}^{n-n_1} a_{k,i} z_i - \sum_{j=1}^{k-1} a_{k,n-n_1+j} z_{n-n_1+j}, \quad (7)$$

где $a_{k,i}$ и b_k — коэффициенты матриц $\mathbf{A}_{1z}$ и $\mathbf{B}_{1z}$ соответственно. Для переменных z_i с порядковым номером, большим чем $n - n_1$, имеем

$$\begin{aligned} z_{n-n_1+1} &= b_1 - \sum_{i=1}^{n-n_1} a_{1,i} z_i; \\ z_{n-n_1+2} &= b_2 - \sum_{i=1}^{n-n_1} a_{2,i} z_i - a_{2,n-n_1+1} z_{n-n_1+1} = \\ &= b_2 - a_{2,n-n_1+1} b_1 - \sum_{i=1}^{n-n_1} [a_{2,i} + a_{2,n-n_1+1} a_{1,i}] z_i; \\ z_{n-n_1+3} &= b_3 - \sum_{i=1}^{n-n_1} a_{3,i} z_i - a_{3,n-n_1+1} z_{n-n_1+1} - \\ &- a_{3,n-n_1+2} z_{n-n_1+2} = b_3 - a_{3,n-n_1+2} b_2 - \end{aligned}$$

$$\begin{aligned} &- [a_{3,n-n_1+1} + a_{3,n-n_1+2} a_{2,n-n_1+1}] b_1 - \\ &- \sum_{i=1}^{n-n_1} \left[a_{3,i} + a_{3,n-n_1+2} a_{2,i} + \right. \\ &\left. + [a_{3,n-n_1+1} + a_{3,n-n_1+2} a_{2,n-n_1+1}] a_{1,i} \right] z_i \text{ и т. д.,} \end{aligned}$$

если в (6) матрица $\mathbf{A}_1$ методом Гаусса и перестановками столбцов приведена к нижней треугольной $\mathbf{A}_{1z}$ вида

$$\mathbf{A}_{1z} \mathbf{z} = \mathbf{B}_{1z}, \quad \mathbf{x} = \mathbf{Tz},$$

$$\mathbf{A}_{1z} = \begin{bmatrix} a_{1,1} & \dots & a_{1,n-n_1} & 1 & 0 & \dots & 0 \\ a_{2,1} & \dots & a_{2,n-n_1} & a_{2,n-n_1+1} & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n_1,1} & \dots & a_{n_1,n-n_1} & a_{n_1,n-n_1+1} & a_{n_1,n-n_1+2} & \dots & 1 \end{bmatrix}, \quad (8)$$

где $\mathbf{z}$ — вектор $\mathbf{x}$ с измененным порядком следования компонентов; $\mathbf{B}_{1z}$ — преобразованный вектор $\mathbf{B}_1$; $\mathbf{T}$ — матрица перестановок. Учесть ограничения (6) можно линейной заменой

$$\mathbf{x} = \mathbf{T} \mathbf{A}_1 \mathbf{z}_{n-n_1} + \mathbf{T} \mathbf{B}_1, \quad (9)$$

где $\mathbf{A}_1$ — матрица размерности $n \times (n - n_1)$, коэффициенты которой находятся согласно (7); $\mathbf{z}_{n-n_1}$ — новый вектор варьируемых переменных; $\mathbf{B}_1$ — вектор размерности $n \times 1$, коэффициенты которого находятся согласно (7).

Построение области кандидатов в решения методом анализа линейных ограничений типа неравенство

Другой широко распространенный вид ограничений — это линейные неравенства, которые задаются выражением вида

$$\mathbf{A}_2 \mathbf{x} \leq \mathbf{B}_2, \quad (10)$$

где матрица $\mathbf{A}_2$ имеет размерности $m_n \times n$; вектор $\mathbf{x}$ имеет размерности $n \times 1$, а $\mathbf{B}_2$ имеет размерности $m_n \times 1$.

Первый шаг: определить взаимоисключающие ограничения, что может быть выполнено по алгоритму 1.

1. Начало алгоритма, ввод исходных данных: матрица $\mathbf{A}_2$ и вектор $\mathbf{B}_2$.
2. Цикл: изменять индекс i от 1 до $m_n - 1$ с шагом 1, пропустить номер из запретного списка: выполнять от п. 3 до п. 12.
3. Переменной флаг1 присвоить 0.
4. Цикл: изменять индекс j от $i + 1$ до m_n с шагом 1, пропустить номер из запретного списка: выполнять от п. 5 до п. 12.
5. Если коэффициенты выбранных неравенств $a_{i,1} \neq 0$ и $a_{j,1} \neq 0$ или выбранные неравенства имеют нулевые коэффициенты от 1 до $s - 1$, то рассчитать коэффициент $k = a_{i,1}/a_{j,1}$ или $k = a_{i,s}/a_{j,s}$, иначе перейти в п. 4.

6. Если для всех p от 1 до n хотя бы одно из $a_{i,p} - a_{j,p}k \neq 0$, то перейти в п. 4.
 7. Если $b_i - b_jk = 0$, то исключить ограничение j , заменить ограничение i на $a_{i,1}x_1 + a_{i,2}x_2 + \dots + a_{i,n}x_n = b_i$, добавить в список исключенных ограничений индекс j , добавить в запретный список индекс j , переменной флаг1 присвоить 1, перейти в п. 4.
 8. Если переменная флаг1 равна 1 и $b_i - b_jk \neq 0$, то противоречие в ограничениях, задача не имеет решения, добавить в список несовместных ограничений индексы i и j , перейти в п. 13.
 9. Если $k > 0$ и $b_i - b_jk > 0$, то исключить ограничение j , добавить в список исключенных ограничений индекс j , добавить в запретный список индекс j , перейти в п. 4, не изменяя значения j .
 10. Если $k > 0$ и $b_i - b_jk < 0$, то заменить ограничение i на j , добавить в список исключенных ограничений индекс i , добавить в запретный список индекс j , перейти в п. 4.
 11. Если $k < 0$ и $b_i - b_jk > 0$, то параллельные ограничения, добавить в список параллельных ограничений индексы i и j , перейти в п. 4.
 12. Если $k < 0$ и $b_i - b_jk < 0$, то противоречие в ограничениях, задача не имеет решения, добавить в список несовместных ограничений индексы i и j , перейти в п. 13.
 13. Окончание алгоритма, выходные данные: список исключенных ограничений, список параллельных ограничений, список несовместных ограничений, задача имеет / не имеет решения, формируется новая система ограничений
- $$\mathbf{A}_2^* \mathbf{x} \leq \mathbf{B}_2^*. \quad (11)$$

Второй шаг: определить границу области, формируемой ограничениями, или убедиться, что такой области не существует. Известные алгоритмы решения подобной задачи [4, 5] адаптируем для нашего случая.

Введем следующие определения и допущение.

1. Каждое ограничение типа неравенство из (11) имеет *граничную плоскость*, которая определяется уравнением $a_{i,1}x_1 + a_{i,2}x_2 + \dots + a_{i,n}x_n - b_i = 0$.

2. *Узлом (вершиной)* будем называть точку пересечения n граничных плоскостей, и эта точ-

ка должна удовлетворять всем ограничениям из (11); узел будем обозначать как $V_{i,j,\dots,s}$, где $i, j, \dots, s$ — код узла, состоящий из n индексов и формируемый номерами ограничений в (11); $V_{i,j,\dots,s}$ — вектор координат узла с размерностью $n \times 1$; индексы упорядочиваются таким образом, чтобы $i < j < \dots < s$.

3. *Ребром* будем называть отрезок, соединяющий два узла с числом общих индексов, равным двум; ребро будем обозначать как $W_{i,j,\dots,s}^{i,j,\dots,s}$, где верхний и нижний индексы — коды узлов, которые соединяет ребро.

4. *Гранью* будем называть область граничной плоскости, которой принадлежат все узлы, для которых один из индексов кодов i является общим, и эта область сформирована ребрами, соединяющими эти узлы; грань будем обозначать как P_i , где i — это номер соответствующего ограничения из (11).

5. Все варьируемые переменные x изменяются в соответствующих интервалах $x_{\min,i} \leq x_i \leq x_{\max,i}$, что обуславливает реализуемость решения задачи; тогда следует, что система не имеет решения, когда нет ни одного узла; в случае одного узла система имеет единственное решение, если оно не противоречит ограничениям в (1); если есть два узла, то область кандидатов в решение — это ребро; если три узла, то область кандидатов в решение — это грань и т. д.

Центральным вопросом определения области кандидатов в решение является отыскание узлов для системы (11), что может быть выполнено по алгоритму 2.

1. Начало алгоритма, ввод исходных данных: матрица $\mathbf{A}_2^*$ и вектор $\mathbf{B}_2^*$.
2. Цикл: пока не выбраны все возможные последовательности из n неповторяющихся индексов $i < j < \dots < s$ такие, что не могут включаться индексы двух взаимно параллельных граничных плоскостей одновременно, выполнять от п. 2 до п. 6.
3. Из выбранных различных n граничных плоскостей с индексами $i < j < \dots < s$ сформировать систему уравнений $\mathbf{A}_{i,j,\dots,s} \mathbf{x} = \mathbf{B}_{i,j,\dots,s}$.
4. Если $\det(\mathbf{A}_{i,j,\dots,s}) = 0$, то перейти в п. 1.
5. Вычислить $V_{i,j,\dots,s} = (\mathbf{A}_{i,j,\dots,s})^{-1} \mathbf{B}_{i,j,\dots,s}$.
6. Если ограничения (11) от $V_{i,j,\dots,s}$ выполняются, то добавить $V_{i,j,\dots,s}$ в список узлов, перейти в п. 2.
7. Если список узлов пустой, то задача не имеет решения.
8. Если число узлов равно 1, то задача имеет единственное решение.
9. Окончание алгоритма, выходные данные: список узлов, задача имеет / не имеет решения.

Если найдены узлы для (11), то исходя из определений 3 и 4 простыми средствами могут быть построены все ребра и грани, образующие область кандидатов в решения.

Анализ полиномиальных ограничений

Другой широко распространенный вид ограничений — это полиномиальные уравнения. Разработано много методов и алгоритмов для их решения [6–8]. Следует отметить, что полиномиальные аппроксимации могут быть использованы для описания с некоторой точностью более сложных нелинейных ограничений в выбранной области аргументов. Это позволяет сводить для некоторых областей значений аргумента задачу условной оптимизации с произвольными нелинейными ограничениями к задачам условной оптимизации с полиномиальными ограничениями при заданной точности решения.

Полиномиальные ограничения встречаются двух видов — полиномиальные равенства и неравенства, которые имеют различные методы в решении. Универсальным подходом в анализе ограничений и первого, и второго вида может считаться разложение на множители порядка не старше двух полинома ограничения φ :

$$\begin{aligned} \varphi(x_1, x_2, \dots, x_n) = & \prod_{i_1} \left(\sum_{j_1} a_{1,i_1,j_1} x_{j_1} + b_{1,i_1} \right) \times \\ & \times \prod_{i_2} \left(\sum_{j_2} a_{2,i_2,j_2} x_{j_2}^2 + \sum_{j_3, j_4} a_{3,i_2,j_3,j_4} x_{j_3} x_{j_4} + \right. \\ & \left. + \sum_{j_5} a_{4,i_2,j_5} x_{j_5} + b_{2,i_2} \right), \end{aligned} \quad (12)$$

где $a_{1,i_1,j_1}, b_{1,i_1}, a_{2,i_2,j_2}, a_{3,i_2,j_3,j_4}, a_{4,i_2,j_5}, b_{2,i_2}$ — коэффициенты разложения на множители.

Видно, что ограничение с полиномом (12) сводится к системе линейных и квадратичных ограничений, что приводит к разделению исходной задачи на несколько подзадач, которые могут решаться независимо.

В общем случае разложение полиномиальных ограничений — это сложная проблема, исследования в этом направлении сейчас идут очень интенсивно.

Рассмотрим случай систем ограничений, приводимых к треугольному виду:

$$\begin{cases} v_1^{m_1} + \varphi_1(v_1, v_2, \dots, v_k) = 0 \\ v_2^{m_2} + \varphi_2(v_2, v_3, \dots, v_k) = 0, \\ \dots \\ v_k^{m_k} + \varphi_k(v_k) = 0 \end{cases}, \quad \mathbf{v} = \mathbf{T}_v \mathbf{x} + \mathbf{B}_v, \mathbf{v} = [v_1, v_2, \dots, v_k]^T, k \leq n, \quad (13)$$

где $m_1, m_2, \dots, m_k$ — степени соответствующих переменных; k — количество уравнений; φ_i — полиномы от переменных $v_i, \dots, v_k$, причем показатели при переменной v_i в многочлене φ_i меньше, чем m_i ; $\mathbf{T}_v, \mathbf{B}_v$ — матрицы параметров.

В последнее уравнение в (13) входит только v_k , и существуют эффективные алгоритмы нахождения корней полинома одной переменной [8]. Рассматриваются только вещественные корни полинома уравнения k , которые могут быть подставлены в уравнение $k-1$, которое после замены станет зависеть от переменной v_{k-1} , и его корни могут быть найдены аналогично уравнению k и т. д. Если все корни хотя бы одного из уравнений в (13) комплексные, то задача (1) не имеет решения. Множество вещественных корней системы (13) формирует наборы линейных ограничений вида

$$\mathbf{T}_v \mathbf{x} + \mathbf{B}_v = \Lambda_{v,i}, \quad (14)$$

где $\Lambda_{v,i}$ — i -й набор вещественных корней системы (13), и разделяет исходную задачу на несколько параллельно решаемых подзадач. Ограничения (14) анализируются аналогично (6).

Для ограничений типа неравенство рассмотрим случай

$$\begin{aligned} a_i v_i^{m_i} + \varphi_i(v_i) & \leq 0, i = 1, 2, \dots, k, \\ \mathbf{v} = \mathbf{T}_v \mathbf{x} + \mathbf{B}_v, \mathbf{v} & = [v_1, v_2, \dots, v_k]^T, \end{aligned} \quad (15)$$

где количество неравенств может быть больше количества переменных, а каждый полином φ_i зависит только от одной переменной v_i . Найдя корни полинома, представим (15) как произведение слагаемых только с вещественными корнями λ_j :

$$a_i \prod_j (v_i - \lambda_j) \leq 0. \quad (16)$$

Из (16) находятся интервалы для v_i , которые преобразуются в набор систем линейных неравенств от $\mathbf{x}$, и необходимо рассмотреть все возможные комбинации этих систем для всех неравенств из (15) и решить их с использованием алгоритмов из предыдущего раздела.

Анализ ограничений с разрывами

Часто встречающийся на практике тип ограничений — это ограничения с разрывами. К ним относятся ограничения в виде кусочно-линейных функций, дробей, полиномов (например, гиперболы) и т. д. Наиболее удобным способом представления таких ограничений является описание в виде систем равенств или неравенств, заданных на интервалах, границами которых являются особые точки, например переход с одной ветви

функции на другую. Применим данный подход для дроби

$$\frac{a}{\prod_{i,j} (v_i - \lambda_j)} \leq 0, \quad (17)$$

где вещественные корни для каждой переменной v_i упорядочены согласно правилу $\lambda_1 < \lambda_2 < \dots < \lambda_p$, тогда (17) можно представить как системы неравенств, соответствующие интервалам λ . Для случая (17) при $a < 0$ необходимо, чтобы выполнялись условия $v_i - \lambda_j > 0$ для всех компонент или количество условий $v_i - \lambda_j < 0$ было четным для выбранных интервалов v_i . Анализ систем неравенств выполняется аналогично (6)–(9) и алгоритмам 1 и 2 при условии, что x , удовлетворяющие уравнениям $v_i - \lambda_j = 0$, не могут быть решениями. Если узел, ребро или грань удовлетворяет этим условиям, то в качестве кандидата в решения необходимо выбрать точку, линию или плоскость, лежащую в области кандидатов в решение и отстоящую на точность расчета ε .

Разбиение на интервалы приводит к возникновению нескольких задач, которые могут решаться независимо, и все их возможные комбинации должны быть рассмотрены. Наиболее простым способом описания подобного разбиения является представление в виде логической структуры, например конечного автомата [9].

Анализ ограничений с разрывами должен выполняться после всех других типов ограничений, так как в этом случае повышается вероятность исключения подобных ограничений или областей аргументов, которые соответствуют разрывам.

Анализ произвольных ограничений

Каждый тип ограничений требует разработки соответствующего алгоритма анализа. Наиболее простым способом решения этой задачи является аппроксимация произвольного нелинейного ограничения кусочно-полиномиальной моделью при заданной точности расчета ε . В этом случае можно воспользоваться подходом, описанным в двух предыдущих разделах.

При синтезе оптимальных систем управления в качестве ограничений используются уравнения объекта управления, компонентами которых являются экспоненциальные функции. Подобный пример подробно описан в работе [10], а ограничения имеют вид

$$x_i e^{-b_i x_1} - a_{1,i} e^{-x_1 x_i} - a_{2,i} x_i + a_{3,i} = 0; \quad (18)$$

$$a_j e^{-b_j x_j} - x_i e^{-x_i x_j} = 0; \quad (19)$$

$$\frac{e^{-b_i x_i} - e^{-x_i x_j}}{\frac{a_{1,j}}{x_i} - a_{2,j}} \leq c_i, \quad (20)$$

где $b_i, a_{1,i}, a_{2,i}, a_{3,i}, a_j, b_j, a_{1,j}, a_{2,j}, c_i$ — постоянные коэффициенты, $i \neq j$.

Нетрудно заметить в (18), что переменные x_i могут быть выражены через x_1 , для этого введем замену $y_i = x_1 x_i$, тогда (18) примет вид

$$e^{-y_i} = -\mu_i y_i + \sigma_i, \quad \mu_i(x_1) = \frac{a_{2,i} - e^{-b_i x_1}}{a_{1,i} x_1},$$

$$\sigma_i = \frac{a_{3,i}}{a_{1,i}}. \quad (21)$$

Уравнение (21) имеет два одинаковых решения в случае

$$\frac{\partial e^{-y_i}}{\partial y_i} = -e^{-y_i} = \frac{\partial(-\mu_i y_i + \sigma_i)}{\partial y_i} = -\mu_i, \quad (22)$$

что позволяет уточнить нижнюю границу для σ_i :

$$y_i = -\ln \mu_i, \quad \mu_i = \mu_i \ln \mu_i + \sigma_i, \quad \sigma_i = \mu_i - \mu_i \ln \mu_i. \quad (23)$$

Учтем (23) в уравнении (21):

$$e^{-y_i} = -\mu_i y_i + \sigma_i, \quad \mu_i > 0, \quad \mu_i - \mu_i \ln \mu_i - \sigma_i < 0. \quad (24)$$

Перегруппировав (24): $-\frac{1}{\mu_i} = \left(y_i - \frac{\sigma_i}{\mu_i}\right) e^{y_i}$, и введя замену $z_i = y_i - \frac{\sigma_i}{\mu_i}$, получим

$$z_i e^{z_i} = -\frac{1}{\mu_i} e^{\frac{\sigma_i}{\mu_i}}. \quad (25)$$

Решением (25) является функция Ламберта ψ^{-1} :

$$z_i = f^{-1} \left(-\frac{1}{\mu_i} e^{\frac{\sigma_i}{\mu_i}} \right),$$

$$y_i = \frac{\sigma_i}{\mu_i} + z_i = \frac{\sigma_i}{\mu_i} + \psi^{-1} \left(-\frac{1}{\mu_i} e^{\frac{\sigma_i}{\mu_i}} \right). \quad (26)$$

Тогда переменные x_i находятся из соотношения

$$x_i = \frac{\sigma_i}{\mu_i(x_1) x_1} + \frac{1}{x_1} \psi^{-1} \left(-\frac{1}{\mu_i(x_1)} e^{\frac{\sigma_i}{\mu_i(x_1)}} \right) = g_i(x_1),$$

$$\mu_i(x_1) > 0, \quad \mu_i(x_1) - \mu_i(x_1) \ln \mu_i(x_1) - \sigma_i < 0. \quad (27)$$

Нетрудно заметить, что переменные x_j могут быть выражены через $g(x_1)$ с использованием ограничений (19):

$$\frac{a_j}{g_i(x_1)} = e^{(b_j - g_i(x_1))x_j} \Rightarrow$$

$$\Rightarrow x_j = \frac{1}{b_j - g_i(x_1)} \ln \left(\frac{a_j}{g_i(x_1)} \right) = h_i(x_1), \quad (28)$$

$$\frac{a_j}{g_i(x_1)} > 0.$$

Тогда с учетом (28) ограничения (20) будут зависеть только от x_1 и примут вид

$$\frac{e^{-b_i g_i(x_1)} - e^{-g_i(x_1) h_i(x_1)}}{\frac{a_{1,j}}{g_i(x_1)} - a_{2,j}} \leq c_i. \quad (29)$$

Если учесть выражения (21)–(29), то ограничения (18)–(20) сводятся к эквивалентной системе ограничений от переменной x_1 :

$$\begin{aligned} \mu_i(x_1) > 0, \mu_i(x_1) - \mu_i(x_1) \ln \mu_i(x_1) - \sigma_i < 0, \\ x_1 \neq 0, g_i(x_1) \neq 0, b_j - g_i(x_1) \neq 0, \\ \frac{e^{-b_i g_i(x_1)} - e^{-g_i(x_1) h_i(x_1)}}{\frac{a_{1,j}}{g_i(x_1)} - a_{2,j}} \leq c_i, \frac{a_j}{g_i(x_1)} > 0. \end{aligned} \quad (30)$$

Нелинейные функции в (30) могут быть аппроксимированы с помощью кусочно-полиномиальной модели от одной переменной x_1 и проанализированы, как это рассмотрено выше.

Анализ целевой функции

После построения границы области допустимых решений и определения ее узлов путем анализа ограничений необходимо определить оптимальные значения целевой функции, которые принадлежат области допустимых решений. На основании примера, описанного в первом разделе, сформулируем общий вывод: если решается задача условной минимизации (максимизации), целевая функция гладкая и ее экстремумы могут быть определены аналитически, тогда возможны три варианта.

1. Если глобальный минимум (максимум) находится в области допустимых значений, образованной ограничениями, тогда аргумент, соответствующий этому минимуму (максимуму), является решением поставленной задачи.

2. Если область допустимых решений — это пустое множество, тогда поставленная задача не имеет решения.

3. Если глобальный минимум (максимум) находится вне области допустимых значений, образованной ограничениями, тогда решение задачи будет находиться на границе области допустимых решений или это будут аргументы локальных минимумов (максимумов), которые находятся внутри области допустимых решений. В данном случае задача может иметь несколько решений, которые будут соответствовать одинаковым минимальным (максимальным) значениям целевой функции.

Обычно целевая функция имеет простой вид и ее оптимумы известны, поэтому проверка вариантов 1 и 2 не вызывает трудностей. Для про-

извольной целевой функции проверка варианта 3 очень трудоемка даже на небольшой области допустимых решений и может быть проведена стандартными методами поиска, например методом сеток, методом управляемого случайного поиска.

Рассмотрим случай, когда целевая функция является метрикой; имеет не более одного экстремума; все точки, в которых эта функция имеет одинаковое значение, ограничивают фигуру таким образом, что отрезок, соединяющий любые две точки ее поверхности, принадлежит этой фигуре; выполняются следующие соотношения.

Если определена функция ρ такая, что $\rho = \rho(\mathbf{x})$, и существуют такие точки $\mathbf{x}(i)$, в которых $\rho(\mathbf{x}(i))$ имеют одинаковые значения, то должны выполняться следующие условия.

Если ищется минимум функции f и эта функция имеет максимум в точке, где $\rho = \rho_0$, то

$$\begin{aligned} f(\rho_0) = \max, f(\rho_0) > f(\rho_1) > \dots > f(\rho_n), \\ \rho_0 < \rho_1 < \dots < \rho_n; \end{aligned} \quad (31a)$$

если ищется максимум функции f и эта функция имеет минимум в точке, где $\rho = \rho_0$, то

$$\begin{aligned} f(\rho_0) = \min, f(\rho_0) < f(\rho_1) < \dots < f(\rho_n), \\ \rho_0 > \rho_1 > \dots > \rho_n. \end{aligned} \quad (31b)$$

Тогда решением будут узлы, в которых целевая функция имеет максимальное/минимальное значение для задачи максимизации/минимизации.

Если ищется минимум функции f и эта функция имеет минимум в точке, где $\rho = \rho_0$, то

$$\begin{aligned} f(\rho_0) = \min, f(\rho_0) < f(\rho_1) < \dots < f(\rho_n), \\ \rho_0 < \rho_1 < \dots < \rho_n; \end{aligned} \quad (32a)$$

если ищется максимум функции f и эта функция имеет максимум в точке, где $\rho = \rho_0$, то

$$\begin{aligned} f(\rho_0) = \max, f(\rho_0) > f(\rho_1) > \dots > f(\rho_n), \\ \rho_0 > \rho_1 > \dots > \rho_n. \end{aligned} \quad (32b)$$

Тогда решением будут узлы или точки граней, если эти грани являются касательными к поверхности ρ в этих точках, где целевая функция имеет максимальное/минимальное значение для задачи максимизации/минимизации.

Алгоритм оптимизации

Укрупненно алгоритм оптимизации может быть описан следующим образом.

1. Начало алгоритма, ввод данных об ограничениях и целевой функции.

2. Если выделены линейные равенства, то выполнить анализ линейных неравенств согласно (6)–(9).
3. Если не найдены противоречия в ограничениях в п. 2 и выделены линейные неравенства, то выполнить анализ линейных неравенств согласно алгоритмам 1 и 2.
4. Если не найдены противоречия в ограничениях в п. 3 и выделены полиномиальные ограничения, то
выполнить анализ полиномиальных ограничений согласно (12)–(16).
5. Если не найдены противоречия в ограничениях в п. 4 и выделены ограничения с разрывами, то
выполнить анализ полиномиальных ограничений согласно разделу «Анализ ограничений с разрывами».
6. Если не найдены противоречия в ограничениях в п. 5 и выделены произвольные ограничения, то
выполнить анализ ограничений согласно разделу «Анализ произвольных ограничений».
7. Если найдены противоречия в ограничениях, то
задача не имеет решения,
перейти в п. 12.
8. Найти все узлы области допустимых решений.
9. Вычислить значения целевой функции в узлах – найти ее максимальные/минимальные значения.
10. Если целевая функция удовлетворяет условиям (31a) и (31b), то
исходя из задачи максимизации/минимизации выбрать узлы с максимальными/минимальными значениями, полученными в п. 9, перейти в п. 12.
11. Если целевая функция удовлетворяет условиям (32a) и (32b), то
исходя из задачи максимизации/минимизации выбрать узлы с максимальными/минимальными значениями, полученными в п. 9, проверить, являются ли выбранные узлы точками касания поверхности p граней области допустимых значений, вычислить точки касаний граней, образующих узел с максимальным/минимальным значением, выбрать точки с максимальным/минимальным значением целевой функции.
12. Окончание алгоритма, выходные данные:
список решений,
задача имеет / не имеет решения.

Заключение

В статье рассмотрены вопросы решения задач условной оптимизации большой размерности, которые часто встречаются в областях науки и техники, таких как теория управления, распознавание образов, моделирование и т. д.

Получены следующие результаты.

1. Показано, что переход от условной оптимизации к безусловной в случае ограничений типа неравенство не тождественен.

2. Сформулированы необходимые и достаточные условия оптимальности для задач условной оптимизации с целевыми функциями в виде метрики, имеющими единственный оптимум.

3. Рассмотрен вопрос построения областей допустимых решений для ограничений произвольного вида путем их аппроксимации кусочно-полиномиальными функциями или рациональными дробями, что позволяет свести исходную задачу к набору взаимосвязанных подзадач с линейными ограничениями, формирующими выпуклые многомерные оболочки, что существенно сужает область поиска.

4. Разработан новый метод нелинейного программирования, позволяющий решать задачи высокой размерности с большим количеством ограничений различных типов. Предложенный метод при нахождении оптимальных значений не использует операцию дифференцирования, что позволяет применять его при анализе стохастических и интервальных данных.

Литература

1. Tabak D., Kuo B. C. Optimal Control by Mathematical Programming. — Englewood Cliffs, New Jersey: Prentice-Hall, 1971. — 280 p.
2. Zangwill W. I. Nonlinear Programming. A Unified Approach. — Englewood Cliffs, New Jersey: Prentice-Hall, 1969. — 312 p.
3. Растринин Л. А. Случайный поиск в задачах оптимизации многопараметрических систем. — Рига: Зинатне, 1965. — 212 с.
4. Motzkin T., Raiffa H., Thompson G., Thrall R. M. The Double Description Method // Contributions to the Theory of Games. — Princeton: Princeton University Press, 1953. — P. 51–73.
5. Шевченко В. Н., Груздев Д. В. Модификация алгоритма Фурье — Моцкина для построения триангуляции // Дискретный анализ и исследование операций. 2003. Т. 10. № 1. С. 53–64.
6. Computer Algebra: Symbolic and Algebraic Computation / Ed. by B. Buchberger, G. Collins, R. Loos. — N. Y.: Springer-Verlag Wien, 1983. — 291 p.
7. Davenport J., Siret Y., Tournier E. Computer Algebra: Systems and Algorithms for Algebraic Computation. — London, Academic Press, 1993. — 313 p.
8. Cox D., Little J., O'Shea D. Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra. — Springer, 2006. — 551 p.
9. Дубаренко В. В. Принципы логического управления динамическими объектами // Информационно-управляющие системы. 2003. № 5(6). С. 2–11.
10. Артеменко Ю. Н., Агапов В. А., Дубаренко В. В., Кучмин А. Ю. Групповое управление актуаторами контррефлектора радиотелескопа // Информационно-управляющие системы. 2012. № 4(59). С. 2–9.

UDC 681.5

doi:10.15217/issn1684-8853.2016.2.2

A Nonlinear Programming Method with Arbitrary RestrictionsKuchmin A. Yu.^a, PhD, Tech., Senior Researcher, radiotelescope@yandex.ru^aInstitute of Problems of Mechanical Engineering of RAS, 61, Bol'shoi Pr. V. O., 199178, Saint-Petersburg, Russian Federation

Introduction: Modeling and identification of nonlinear and non-stationary systems, as well as the analysis and filtration of experimental data often require the use of optimization methods, in particular, nonlinear programming. These problems belong to the class of problems of multidimensional optimization with restrictions. Unfortunately, at the moment there are no efficient numerical methods to solve such problems for the case when the number of restrictions m exceeds manifold the number of the varied parameters n .

Purpose: The goal is to develop new methods and algorithms for multiple-parameter optimization with restrictions. **Results:** A new nonlinear programming method is developed which is called "Multidimensional Patchwork Shell Method" (MPSM). It allows you to solve problems of conditional multiple-parameter optimization for the case of $n \ll m$. The method uses both necessary and sufficient optimality conditions for metric-like criterion functions. It has been shown that if the only minimum (maximum) of a metric is out of the area of admissible values formed by the restrictions, the solution of the problem will be found on the border of the area of admissible decisions. In this case, the problem can have several solutions which will correspond to identical minimum (maximum) values of the criterion function, and all these solutions can be obtained by MPSM. The basis of MPSM is creating an area of admissible decisions by analyzing restrictions of any type and approximating them with piecewise and polynomial restrictions or with rational fractions. This reduces the initial problem to a set of interconnected subtasks with the linear restrictions forming multidimensional shells, significantly narrowing the search area. **Practical relevance:** The proposed method of nonlinear programming has been tested. It will find application in conditional optimization in such fields as physical metrology, optimal control systems and logical-interval systems. This algorithm can be improved by developing more efficient algorithms for new types of restrictions.

Keywords — Nonlinear Programming, Multiple-Parameter Optimization.

References

1. Tabak D., Kuo B. C. *Optimal Control by Mathematical Programming*. Englewood Cliffs, New Jersey, Prentice-Hall, 1971. 280 p.
2. Zangwill W. I. *Nonlinear Programming. A Unified Approach*. Englewood Cliffs, New Jersey, Prentice-Hall, 1969. 312 p.
3. Rastrigin L. A. *Sluchainyi poisk v zadachakh optimizatsii mnogoparametricheskikh sistem* [Random Search in Problems of Optimization of Multiple Parameter Systems]. Riga, Zinatne Publ., 1965. 212 p. (In Russian).
4. Motzkin T., Raiffa H., Thompson G., Thrall R. M. The Double Description Method. In: *Contributions to the Theory of Games*. Princeton, Princeton University Press, 1953. Pp. 51–73.
5. Shevchenko V. N., Gruzdev D. V. Modification of Algorithm of Fourier — Motzkin for Creation of a Triangulation. *Diskretnyi analiz i issledovanie operatsii*, 2003, vol. 10, no. 1, pp. 53–64 (In Russian).
6. *Computer Algebra: Symbolic and Algebraic Computation*. Ed. by B. Buchberger, G. Collins, R. Loos. New York, Springer-Verlag Wien, 1983. 291 p.
7. Davenport J., Siret Y., Tournier E. *Computer Algebra: Systems and Algorithms for Algebraic Computation*. London, Academic Press, 1993. 313 p.
8. Cox D., Little J., O'Shea D. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, 2006. 551 p.
9. Dubarenko V. V. Principles of Logic Automatic Control by Dynamic Objects. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2003, no. 5(6), pp. 2–11 (In Russian).
10. Artemenko Yu. N., Agapov V. A., Dubarenko V. V., Kuchmin A. Yu. Co-operative Control of Subdish Actuators of Radio-Telescope. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2012, no. 4(59), pp. 2–9 (In Russian).

АВТОМАТИЧЕСКОЕ АННОТИРОВАНИЕ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ОДНОРОДНЫХ ТЕКСТОВО-ВИЗУАЛЬНЫХ ГРУПП

А. В. Проскурин^а, аспирант

М. Н. Фаворская^а, доктор техн. наук, профессор

^аСибирский государственный аэрокосмический университет им. академика М. Ф. Решетнёва, Красноярск, РФ

Постановка проблемы: задача автоматического аннотирования изображений нетривиальна: часто обучающие наборы несбалансированы и содержат неполные аннотации, а между визуальными признаками и текстовым описанием изображения наблюдается семантический разрыв. Существующие методы решают эти проблемы, используя для аннотации нового изображения все обучающие изображения и ключевые слова, в том числе заведомо нерелевантные, что потенциально снижает точность и требует лишних вычислений. При этом используются визуальные признаки большой размерности, что также неэффективно в вычислительном плане. В связи с этим возникает необходимость разработки компактного визуального дескриптора и метода аннотирования тестового изображения с помощью небольшой группы наиболее информативных обучающих изображений. **Результаты:** разработана методика автоматического аннотирования изображений, основанная на поиске апостериорной вероятности ассоциации ключевого слова с визуальным дескриптором изображения. Получены шесть глобальных дескрипторов, объединенных в один дескриптор, размер которого уменьшен с помощью метода главных компонент до нескольких сотен элементов. Проведенные экспериментальные исследования показали улучшение точности аннотирования на 7 % и отклика на 1 %. **Практическая значимость:** разработанный компактный визуальный дескриптор и метод автоматического аннотирования изображений на основе формирования однородных текстово-визуальных групп может быть использован в информационно-поисковых системах в сети Интернет для повышения эффективности поиска изображений.

Ключевые слова — автоматическое аннотирование изображений, глобальный визуальный дескриптор, текстово-визуальные группы.

Введение

Поиск изображений в сети Интернет является распространенной функцией, реализация которой в значительной степени полагается на наличие текстового описания. Однако из-за стремительного роста количества изображений аннотирование вручную становится невозможным, а описания, полученные из текста, окружающего изображение на веб-странице, не всегда соответствуют действительности. В связи с этим становится актуальной разработка систем автоматического аннотирования изображений (ААИ), в которых на основе большого количества заранее проаннотированных изображений (обучающих изображений) определяется семантическая модель, автоматически присваивающая новому изображению текстовое описание в виде ключевых слов.

В последние десятилетия были предложены различные подходы к формированию ААИ, наиболее эффективный из которых основан на использовании метода ближайших соседей [1–4]. С его помощью для тестового изображения определяются визуально похожие обучающие изображения (ближайшие соседи), после чего аннотация генерируется путем перемещения ключевых слов от ближайших соседей к аннотируемому изображению. Однако при этом часто наблюдается проблема семантического разрыва — отсутствие

связи между визуальными признаками изображения и его интерпретацией человеком. Также на точность аннотирования сильно влияет несбалансированность обучающей выборки (огромная разница в частоте встречаемости разных ключевых слов) и наличие неполных аннотаций (изображения аннотированы не всеми релевантными ключевыми словами).

Для решения этих проблем был предложен [3] двухпроходный метод *K* ближайших соседей (2PKNN — 2-Pass K-Nearest Neighbor), в котором для каждого ключевого слова формировалась группа обучающих изображений, соответствующих этому ключевому слову. Из каждой группы выбиралось одинаковое количество наиболее похожих изображений, после чего они использовались для формирования аннотации. Данный метод демонстрирует существенное улучшение полноты аннотирования (изображение описывается с помощью большего количества разных релевантных ключевых слов), однако использование заведомо нерелевантных слов снижает общую точность аннотирования. Также в работе [3] предложен алгоритм вычисления весов, отражающих значимость разных визуальных дескрипторов при сравнении двух изображений. Этот алгоритм повышает эффективность работы метода 2PKNN, однако используемые глобальные признаки требуют значительных вычислительных затрат.

В связи с этим в данной статье предлагается расширенный метод 2PKNN, основанный на однородных текстово-визуальных группах (ОТВ-группах; 2PKNN-HTVG — 2PKNN based on Homogeneous Textual-Visual Groups), в котором точность и быстродействие метода 2PKNN улучшается с помощью предварительного разделения обучающих изображений на ОТВ-группы, лишь небольшое число которых используется для аннотирования тестового изображения. Также для описания изображения предложен компактный глобальный визуальный дескриптор, описывающий характеристики как сцены, так и объектов на изображении.

Математическая модель автоматического аннотирования изображений

Любой метод ААИ предполагает наличие обучающего набора TS , состоящего из изображений и соответствующих им текстовых описаний. Пусть $J = \{I_1, \dots, I_M\}$ — коллекция изображений, а $K = \{k_1, \dots, k_N\}$ — словарь, состоящий из N ключевых слов, тогда обучающий набор $TS = \{(I_1, K_1), \dots, (I_M, K_M)\}$, где $K_M \subseteq K$. Предположим, что обучающий набор разделен на несколько непересекающихся ОТВ-групп $H = \{H_1, \dots, H_L\}$, а выбор ключевых слов в процессе аннотирования тестового изображения A зависит от ассоциации изображения с той или иной группой. Обозначим вероятность ассоциации изображения A с ОТВ-группой H_l как $P(H_l|A)$. Так же как и в работе [3], введем условную вероятность $P_l(A|k_n)$ оценки распределения визуального дескриптора изображения для ключевого слова k_n внутри группы H_l . В этом случае аннотирование изображения моделируется как проблема поиска апостериорной вероятности:

$$P(k_n|A) = \sum_{H_l \in H} \left(P(H_l|A) \frac{P_l(A|k_n)P_l(k_n)}{P(A)} \right), \quad (1)$$

где $P_l(k_n)$ — априорная вероятность ключевого слова k_n внутри ОТВ-группы H_l . Поскольку вероятность $P(A)$ является константой, то для упрощения в дальнейшем она не будет учитываться.

Наилучшее ключевое слово для тестового изображения A определяется с помощью следующей формулы:

$$k^* = \arg \max_n P(k_n|A). \quad (2)$$

В качестве аннотации используется O ключевых слов с наибольшей вероятностью $P(k_n|A)$. Значение O выбирается равным среднему значению ключевых слов в описании обучающих изображений ОТВ-группы с наибольшим значением $P(H_l|A)$. Таким образом, для аннотирования те-

стового изображения A необходимо оценить вероятности $P(H_l|A)$, $P_l(A|k_n)$ и $P_l(k_n)$. Для этого предлагается метод ААИ, состоящий из двух частей: формирования из обучающего набора ОТВ-групп и использования метода 2PKNN внутри каждой группы.

Формирование однородных текстово-визуальных групп

На первом этапе алгоритма обучения системы ААИ необходимо разделить обучающий набор изображений на ОТВ-группы. Идея заключается в том, что обучающие изображения одной ОТВ-группы формируют контекст для аннотируемого изображения — если изображение отнесено к какой-либо группе, то оно аннотируется из ограниченного набора ключевых слов этой группы. Также предполагается, что тестовое изображение может принадлежать нескольким ОТВ-группам, но их количество ограничивается визуальным сходством. Это позволяет отсеять заведомо нерелевантные ключевые слова, не потеряв релевантных, а также снизить количество обучающих изображений, участвующих в аннотировании. Для этого каждая из ОТВ-групп должна соответствовать двум условиям:

— все изображения одной группы включают «характерные» ключевые слова (ключевые слова, встречающиеся в описании небольшого количества изображений);

— изображения одной группы имеют существенное визуальное сходство. Эта задача решается в два этапа:

- 1) проводится первичное разделение изображений на группы на основе совместной встречаемости ключевых слов в описаниях изображений;
- 2) изображения кластеризуются в автоматически определяемое количество ОТВ-групп с использованием текстово-визуальных дескрипторов.

Первичное разделение обучающих изображений

Для первичного разделения изображений необходимо построить взвешенный орграф $G = (K, E)$, где вершины являются ключевыми словами из словаря K . В этом случае дуга $e_{i,j}$ соединяет ключевые слова k_i и k_j , если одно или больше обучающих изображений одновременно проаннотировано ключевыми словами k_i и k_j . Вес этой дуги $w_{i,j}$ определяется по формуле

$$w_{i,j} = \frac{N(k_i, k_j)}{N(k_i)}, \quad (3)$$

где $N(k_i, k_j)$ — количество обучающих изображений, имеющих в описании ключевые слова k_i и k_j одновременно; $N(k_i)$ — количество обучающих

изображений, проаннотированных ключевым словом k_i .

Полученный оргграф разделяется на группы с помощью быстрого алгоритма [5], показывающего хорошие результаты при небольших вычислительных затратах. Таким образом, ключевые слова, часто встречающиеся совместно или имеющие похожие семантические значения, с большой вероятностью попадут в одну группу. После этого каждое обучающее изображение присоединяется к той группе, ключевые слова которой чаще встречаются в текстовом описании изображения. Подобное разделение обучающей выборки позволяет с минимальными затратами получить инициализацию ОТВ-групп, а также показывает более стабильный результат кластеризации.

Кластеризация обучающих изображений

Для последующей кластеризации обучающей выборки в ОТВ-группы необходимо каждое изображение I_m представить в виде текстово-визуального дескриптора $\mathbf{TV}_m = (\mathbf{T}_m, \mathbf{V}_m)$, где $\mathbf{T}_m = \{t_1, \dots, t_N\}$ — текстовый дескриптор, а $\mathbf{V}_m = \{v_1, \dots, v_Z\}$ — глобальный визуальный дескриптор. Длина текстового дескриптора равна размеру словаря ключевых слов, а его элементы вычисляются с помощью статистической меры TF-IDF (Term Frequency — Inverse Document Frequency):

$$t_n^m = \frac{\delta(k_n \in K_m)}{F(k_n)}, \quad (4)$$

где $\delta(k_n \in K_m)$ обозначает наличие/отсутствие ключевого слова k_n в описании изображения I_m (принимает значения 1 и 0 соответственно); $F(k_n)$ — частота встречаемости ключевого слова k_n в обучающей выборке.

Вычисление визуального дескриптора будет подробно рассмотрено ниже. При сравнении двух изображений I_i и I_j с помощью их текстово-визуальных дескрипторов сходство вычисляется по формуле

$$D(\mathbf{TV}_i, \mathbf{TV}_j) = \alpha D_T(\mathbf{T}_i, \mathbf{T}_j) + (1 - \alpha) \exp(-D_V(\mathbf{V}_i, \mathbf{V}_j)); \quad (5)$$

$$D_T(\mathbf{T}_i, \mathbf{T}_j) = \frac{\sum_{n=0}^N \min(t_n^i, t_n^j)}{\sqrt{\sum_{n=0}^N t_n^i \sum_{n=0}^N t_n^j}}; \quad (6)$$

$$D_V(\mathbf{V}_i, \mathbf{V}_j) = \sqrt{\sum_{z=0}^Z (v_z^i - v_z^j)^2}, \quad (7)$$

где $D_T(\cdot)$ — косинусная метрика для сравнения текстовых дескрипторов; $D_V(\cdot)$ — евклидово рас-

стояние между визуальными дескрипторами; α — эмпирический коэффициент, изменяющийся в пределах $[0, 1]$.

Чем больше значение $D(\mathbf{TV}_i, \mathbf{TV}_j)$, тем более схожи изображения I_i и I_j . Полученные дескрипторы кластеризуются с использованием модификации расширенной самоорганизующейся инкрементальной нейронной сети (ESOINN — Enhanced Self-Organizing Incremental Neural Network) [6], единственный слой которой постепенно подстраивается под структуру входных данных, определяя количество кластеров и их топологию. Модифицированный алгоритм ESOINN включает следующие шаги.

1. Структура нейронной сети инициализируется путем первичного разделения обучающей выборки. Для этого из каждой сформированной группы выбирается по два дескриптора, с помощью которых формируются узлы сети $r_i \in R$. Узлы, принадлежащие одной группе, соединяются связями.

2. На вход сети подается новый текстово-визуальный дескриптор $\mathbf{TV}$.

3. Определяются два ближайших узла сети (победитель и второй победитель) с помощью формулы (5). Если расстояние между входным дескриптором и победителем или вторым победителем больше соответствующих порогов подобия, то входной дескриптор вставляется в сеть как первый узел нового класса, а алгоритм переходит к шагу 2 для получения нового дескриптора.

Поскольку распределение входных данных заранее неизвестно, то порог подобия s_i обновляется для каждого узла в отдельности по формуле

$$s_i = \min_{j \in R_i} D(\mathbf{W}_i, \mathbf{W}_j), \quad (8)$$

где R_i — набор узлов (соседей), соединенных с узлом r_i ; $\mathbf{W}_i$ — вектор весов узла r_i .

В случае если узел не имеет соседей, порог подобия вычисляется с помощью всех узлов сети:

$$s_i = \max_{j \in R \setminus \{i\}} D(\mathbf{W}_i, \mathbf{W}_j). \quad (9)$$

4. «Возраст» (числовой коэффициент, при создании новой связи равный 0) всех связей победителя увеличивается на 1, после чего решается вопрос о необходимости создания новой связи между победителем и вторым победителем.

5. Обновляется суммарная плотность победителя. Плотность узла p_i вычисляется с помощью среднего расстояния $\bar{d}_i$ от узла до его соседей:

$$p_i = \frac{1}{(1 + \bar{d}_i)^2}. \quad (10)$$

Если среднее расстояние от узла до его соседей большое, то количество узлов в этой области

небольшое и плотность будет низкой, и наоборот. В течение одной итерации вычисляется плотность только для победителя. Суммарная плотность узла h_i определяется следующим образом:

$$h_i = \frac{1}{q} \sum_Q \sum_{\lambda} p_i, \quad (11)$$

где q — количество периодов, в которые плотность узла r_i больше 0; Q — количество прошедших периодов обучения (можно вычислить как $Q = M/\lambda$, где M — общее количество входных дескрипторов); λ — число, обозначающее период обучения сети.

6. Счетчик количества побед U_{win} узла-победителя r_{win} увеличивается на 1, а векторы весов победителя и его узлов-соседей обновляются с помощью входного дескриптора следующим образом:

$$\Delta \mathbf{W}_{win} = \frac{1}{U_{win}} (\mathbf{TV} - \mathbf{W}_{win}); \quad (12)$$

$$\Delta \mathbf{W}_j = \frac{1}{100 U_{win}} (\mathbf{TV} - \mathbf{W}_j), j \in R_{win}. \quad (13)$$

7. Удаляются все связи, «возраст» которых превышает заранее установленное значение age_{max} .

8. Если период обучения сети закончен (количество входных дескрипторов кратно периоду сети λ), то существующие кластеры разбиваются на подклассы в целях обнаружения перекрывающихся областей, после чего из нейронной сети удаляются узлы, являющиеся шумами. Такими считаются узлы r_i , имеющие двух или меньше топологических соседей и удовлетворяющие условию следующего вида:

$$h_i < b_o \sum_{j=1}^R \frac{h_j}{R}, \quad (14)$$

где $b_o, o = \{1, 2, 3\}$ — эмпирические коэффициенты, используемые при удалении узлов с двумя топологическими соседями, одним соседом и не имеющих соседей соответственно.

9. Если процесс кластеризации закончен (на вход сети поданы все дескрипторы), то полученные узлы классифицируются по принадлежности к тому или иному кластеру с использованием понятия пути между двумя узлами (узлы r_i и r_j связаны путем, если между ними существует непрерывная цепочка связей).

10. Если ESOINN продолжает работу, то переходим к шагу 2 для получения нового входного дескриптора.

После окончательного формирования структуры нейронной сети необходимо ассоциировать обучающие изображения с полученными кластерами, являющимися «скелетом» ОТВ-групп. Вначале для каждого изображения определяется ближайший узел сети с помощью только текстового дескриптора, после чего этот же процесс повторяется с использованием только визуального дескриптора. В случае когда изображение по текстовому и визуальным дескрипторам ассоциировано с разными кластерами, изображение считается шумовым и исключается из обучающей выборки. Это необходимый шаг, поскольку при выполнении алгоритма аннотирования ассоциация тестового изображения A происходит только при помощи визуальных дескрипторов. Пример одной из ОТВ-групп, сформированной посредством базы изображений IAPR-TC12 [7], представлен на рис. 1.

Процесс оценки вероятности $P(H_l|A)$ из уравнения (1) включает следующие шаги.

1. Определяется расстояние $ds(H_l, A)$ между изображением A и ОТВ-группой H_l . Для этого с помощью уравнения (7) вычисляется расстояние между A и всеми узлами l -го кластера ESOINN и выбирается наименьшее из них.



■ Рис. 1. Пример изображений однородной текстово-визуальной группы

2. Вычисляется диаметр $da(H_{nn})$ ближайшей к изображению A ОТВ-группы H_{best} как максимальное расстояние между любыми двумя обучающими изображениями группы.

3. Оцениваются условные вероятности $P(H_l|A)$ с помощью формулы

$$P(H_l|A) = \begin{cases} \exp(-ds(H_l, A)), & \text{если } ds(H_l, A) \leq da(H_{nn}). \\ 0 & \text{иначе} \end{cases} \quad (15)$$

4. Условные вероятности $P(H_l|A)$ нормализуются таким образом, чтобы их сумма равнялась 1.

Следует отметить, что эффективность предложенного метода может быть повышена, если вместе с тестовым изображением будут предоставлены некоторые ключевые слова, полученные от пользователей. В этом случае расстояние между новым изображением и ОТВ-группой вычисляется с помощью текстово-визуального дескриптора и уравнения (5).

После разделения обучающего набора на ОТВ-группы каждая из них используется в качестве исходных данных для метода 2PKNN. Пусть J — набор обучающих изображений ОТВ-группы H_l , а $J_n \subseteq J$, $n \in \{1, \dots, N\}$ — набор, содержащий все изображения группы, имеющие в описании ключевое слово k_n . Поскольку изображения набора J_n включают одно общее ключевое слово, будем называть такой набор семантической группой. Так как изображение обычно проаннотировано несколькими ключевыми словами, то оно может принадлежать нескольким семантическим группам. Следует отметить, что ОТВ-группы имеют ограниченный набор ключевых слов и, таким образом, некоторые семантические группы могут быть пустыми.

При аннотировании тестового изображения A из каждой семантической группы J_n с помощью уравнения (7) выбирается Y наиболее похожих изображений, формирующих набор $J_{A,n}$. Таким образом, каждый набор $J_{A,n}$ содержит изображения, наиболее информативные при оценке вероятности принадлежности ключевого слова k_n тестовому изображению A . В связи с этим в оценке вероятности $P_l(A|k_n)$ участвуют только изображения из набора $J_{A,n}$:

$$P_l(A|k_n) = \sum_{I_i \in J_{A,n}} \exp(-D_V(A, I_i)). \quad (16)$$

Полученные условные вероятности нормализуются для того, чтобы их сумма равнялась 1. Поскольку для оценки вероятности каждого ключевого слова используется одинаковое количество изображений, то априорная вероятность $P_l(k_n)$ в уравнении (1) одинакова для всех ключевых слов:

$$P_l(k_n) = \frac{1}{N(H_l)}, \quad (17)$$

где $N(H_l)$ — количество уникальных ключевых слов в текстовом описании изображений ОТВ-группы H_l .

Вычисление глобального визуального дескриптора

В эффективности работы предложенного метода ААИ большое значение имеет точность представления изображений в виде визуальных дескрипторов. Наиболее успешные подходы, предложенные для решения этой проблемы, включают три шага: извлечение из изображений локальных признаков (таких как SIFT [8], SURF [9] и т. д.); формирование словаря визуальных слов; кодирование локальных признаков для формирования глобального дескриптора (например, методами SC [10], LLC [11], VLAD [12]). Рассмотрим их подробнее.

1. На первом этапе изображения описываются с помощью набора локальных признаков $X = \{x_1, \dots, x_B\}$, где $x_i \in \mathbb{R}^D$. В работе [13] для классификации изображений был предложен метод быстрого вычисления локальных признаков FDG-SUF, состоящий из двух этапов: вычисления матрицы частей локальных дескрипторов DS и построения с ее помощью набора локальных дескрипторов. На первом этапе все изображение разделяется сеткой на ячейки размером 5×5 пикселей. После этого в каждой из ячеек вычисляются части дескриптора G-SURF [14], которые сохраняются в матрицу DS . На втором этапе по матрице DS перемещается скользящее окно размером 4×4 ячейки. Каждый локальный дескриптор представляет собой объединение частей дескриптора, попавших в скользящее окно. Таким образом, изменяя шаг смещения скользящего окна, можно существенно увеличить количество локальных дескрипторов, извлеченных из изображения, без значительных вычислительных затрат.

2. На следующем шаге формируется словарь визуальных слов $VW = \{vw_1, \dots, vw_S\}$, где $vw_i \in \mathbb{R}^D$. Для этого с помощью алгоритма k -средних выбранные случайным образом локальные признаки кластеризуются. Количество кластеров обычно устанавливается в пределах от 16 до 256. Центр масс кластера выбирается в качестве визуального слова vw_i .

3. По сформированному словарю локальные признаки кодируются в один глобальный вектор $C \in \mathbb{R}^{S \times D}$ с помощью алгоритма VLAD (Vector of Locally Aggregated Descriptors) [12]. Суть метода заключается в том, что для каждого локального признака x_i находится ближайшее визуальное слово $NN(x_i)$. После этого для каждого визуального слова накапливается разница c_j между ним и ассоциированными с ним локальными

признаками. В отличие от оригинального алгоритма, в котором для вычисления разницы используется выражение $NN(\mathbf{x}_i) - \mathbf{x}_i$, в этой работе вклад каждого локального вектора уравнивается:

$$\mathbf{c}_j = \sum_{\mathbf{x}_i: NN(\mathbf{x}_i) = \mathbf{vw}_j} \frac{\mathbf{vw}_j - \mathbf{x}_i}{\|\mathbf{vw}_j - \mathbf{x}_i\|_2}. \quad (18)$$

После вычислений всех $\mathbf{c}_j$ они нормализуются с помощью L2-нормы и объединяются, формируя глобальный дескриптор размером $S \times D$.

Полученный дескриптор показывает хорошие результаты при классификации изображений по типу сцены. Для вычисления дескриптора, описывающего объекты на изображении, используется такой же алгоритм, однако на первом этапе локальные признаки извлекаются только в особых точках, полученных с помощью матриц Гессе [14].

В работе [15] было показано, что некоторые локальные цветовые дескрипторы, имеющие высокую инвариантность к изменениям интенсивности цветов, могут повысить точность классификации изображений. В этой работе в качестве локальных дескрипторов используются G-SURF, OppG-SURF и RGBG-SURF, вычисленные на изображениях в оттенках серого, цветовых пространствах Opponent и нормализованном RGB соответственно.

Таким образом, каждое изображение описывается с помощью шести глобальных визуальных дескрипторов. Для снижения дальнейших вычислительных затрат все дескрипторы объединяются в один, после чего его размерность сокращается по методу главных компонент (PCA) [16].

Результаты экспериментальных исследований

Для экспериментов использовалась база изображений IAPR TC-12 [7], содержащая 19 627 изображений размером 480×360 пикселей, каждое из которых описано несколькими предложениями. В работе [1] для базы был предложен словарь из 291 ключевого слова, состоящий из наиболее часто встречающихся существительных. Для обучения используется 17 665 изображений, остальные 1962 изображения применяются для тестирования.

С помощью этих изображений проводилось сравнение предложенного метода 2PKNN-HTVG с существующими методами ААИ. Оценка эффективности заключалась в вычислении четырех параметров: средней точности (precision), среднего отклика (recall), F1-меры и количества ключевых слов с положительным откликом (N+):

$$\text{precision} = \frac{1}{N} \sum_{n=1}^N \frac{CA(k_n)}{AA(k_n)}, \quad (19)$$

$$\text{recall} = \frac{1}{N} \sum_{n=1}^N \frac{CA(k_n)}{GT(k_n)}, \quad (20)$$

$$F1 = 2 \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \quad (21)$$

где $AA(k_n)$ — количество изображений, автоматически аннотированных ключевым словом k_n ; $CA(k_n)$ — количество изображений, правильно аннотированных ключевым словом k_n ; $GT(k_n)$ — количество изображений, содержащих в тестовой аннотации ключевое слово k_n .

Все вычисления повторялись 5 раз, после чего выбирался лучший результат. Для работы метода 2PKNN-HTVG на разных этапах были установлены следующие параметры.

1. Для формирования каждого словаря визуальных слов из обучающего набора случайным образом выбиралось 200 000 локальных признаков, которые с помощью алгоритма k -средних кластеризовались в 128 кластеров (визуальных слов).

2. Объединенный глобальный дескриптор уменьшался с помощью метода PCA до 256 элементов.

3. При формировании ОТВ-групп текстовый дескриптор изображения имел большее значение ($\alpha = 0,75$), а параметры сети ESOINN устанавливались следующими: $\lambda = 50$; $age_{\max} = 25$; $b_1 = 0,0$; $b_2 = 0,1$; $b_3 = 1,0$.

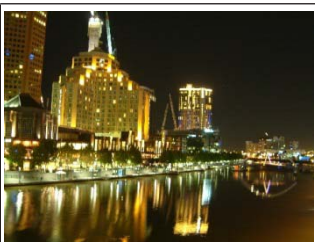
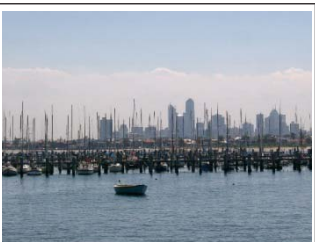
4. Количество изображений Y , выбираемых из семантических групп в методе 2PKNN, равнялось двум.

Примеры аннотирования некоторых изображений показаны на рис. 2, а–в.

В таблице приведены полученные числовые оценки эффективности предложенного метода 2PKNN-HTVG. Оценки для существующих методов ААИ взяты из соответствующих статей.

■ Оценка эффективности разных методов ААИ

Метод	Точность, %	Отклик, %	F1-мера	N+
MBRM [17]	24	23	23,5	223
JEC [1]	28	29	28,5	250
GS [18]	32	29	30,4	252
TagProp(ML) [2]	48	25	32,9	227
TagProp(σML) [2]	46	35	39,8	266
FastTag [4]	47	26	33,5	280
2PKNN [3]	49	32	38,7	274
2PKNN-ML [3]	54	37	43,9	278
2PKNN-HTVG	61	38	46,8	271

a)				
б)	bush, front, rock, sign, wood	building, night, reflection, river, tree, water	boat, building, city, cloud, sky, skyline	board, car, fence, grey, people, racetrack, racing, spectator
в)	bush, sign, rock, wood, front	building, night, river, tree, reflection	boat, building, city, cloud, sky	board, car, fence, people, racetrack, racing, spectator

■ **Рис. 2.** Примеры изображений из базы IAPR-TC12: *а* — исходные изображения; *б* — достоверные аннотации изображений; *в* — аннотации, полученные с помощью метода 2PKNN-HTVG

Анализ результатов, приведенных в таблице, показывает, что предложенный метод эффективнее оригинального алгоритма 2PKNN по точности аннотирования на 7 %. При этом отклик увеличился на 1 % за счет более точного подбора количества ключевых слов в описании аннотируемых изображений.

Заключение

В статье представлен метод автоматического аннотирования изображений, основанный на разделении обучающего набора изображений на ОТВ-группы и аннотировании нового изображения с помощью обучающих изображений небольшого

количества визуально похожих ОТВ-групп. Это позволяет при аннотировании сузить поиск наиболее информативных обучающих изображений и тем самым повысить быстродействие и точность аннотирования. Также представлен алгоритм вычисления компактного глобального визуального дескриптора, описывающего как сцену, так и объекты на изображении. Проведенные экспериментальные исследования показали, что использование предложенного метода повышает точность аннотирования на 7 %, а более точный подбор количества ключевых слов в описании увеличивает отклик на 1 %. Следует отметить, что качество аннотирования может быть повышено с помощью предоставленных пользователем неполных аннотаций.

Литература

1. Makadia A., Pavlovic V., Kumar S. A New Baseline for Image Annotation // Proc. 10th European Conf. on Computer Vision, Marseille, France, 2008. Vol. 5304. P. 316–329.
2. Guillaumin M., Mensink T., Verbeek J., Schmid C. TagProp: Discriminative Metric Learning in Nearest Neighbor Models for Image Auto-Annotation // Proc. IEEE 12th Intern. Conf. on Computer Vision, Kyoto, Japan, 2009. P. 309–316.
3. Verma Y., Jawahar C. V. Image Annotation Using Metric Learning in Semantic Neighbourhoods // Proc. 12th European Conf. on Computer Vision, Florence, Italy, 2012. Vol. 7574. P. 836–849.
4. Chen M., Zheng A., Weinberger K. Q. Fast Image Tagging // Proc. 30th Intern. Conf. on Machine Learning, Atlanta, USA, 2013. P. 1274–1282.
5. Blondel V. D., Guillaume J. L., Lambiotte R., Lefebvre E. Fast Unfolding of Communities in Large Networks // Journal of Statistical Mechanics: Theory and Experiment. 2008. Vol. 2008. P10008.
6. Shen F., Ogura T., Hasegawa O. An Enhanced Self-Organizing Incremental Neural Network for Online Unsupervised Learning // Neural Networks. 2007. Vol. 20(8). P. 893–903.
7. IAPR TC-12 Benchmark. <http://www-i6.informatik.rwth-aachen.de/imageclef/resources/iaprtc12.tgz> (дата обращения: 22.02.2016).
8. Lowe D. G. Distinctive Image Features from Scale-Invariant Keypoints // Intern. Journal of Computer Vision. 2004. Vol. 60(2). P. 91–110.
9. Bay H., Ess A., Tuytelaars T., Gool L. V. Speeded-Up Robust Features (SURF) // Computer Vision and Image Understanding. 2008. Vol. 110(3). P. 346–359.
10. Yang J., Yu K., Gong Y., Huang T. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification // Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Miami, USA, 2009. P. 1794–1801.
11. Wang J., et al. Locality-Constrained Linear Coding for Image Classification/ J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong // Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Francisco, USA, 2010. P. 3360–3367.

12. Jegou H., Douze M., Schmid C., Perez P. Aggregating Local Descriptors into a Compact Image Representation // *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, USA, 2010. P. 3304–3311.
13. Проскурин А. В. Быстрый локальный дескриптор для категоризации изображений по типу сцены // Решетневские чтения: материалы XIX Междунар. науч.-практ. конф., Красноярск, 10–14 ноября 2015 г. Красноярск, 2015. Т. 2. С. 243–245.
14. Alcantarilla P. F., Bergasa L. M., Davison A. J. Gauge-SURF Descriptors // *Image and Vision Computing*. 2013. Vol. 31(1). P. 103–116.
15. Favorskaya M., Proskurin A. Image Categorization Using Color G-SURF Invariant to Light Intensity // *Procedia Computer Science*. 2015. Vol. 60. P. 681–690.
16. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерности. — М.: Финансы и статистика, 1989. — 607 с.
17. Feng S., Manmatha R., Lavrenko V. Multiple Bernoulli Relevance Models for Image and Video Annotation // *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Washington, USA, 2004. Vol. 2. P. 1002–1009.
18. Zhang S., Huang J., Li H., Metaxas D. N. Automatic Image Annotation and Retrieval Using Group Sparsity // *IEEE Transactions on Systems, Man, and Cybernetics*. Part B: Cybernetics. 2012. Vol. 42(3). P. 838–849.

UDC 004.932.2

doi:10.15217/issn1684-8853.2016.2.11

Automatic Image Annotation Based on Homogeneous Textual-Visual GroupsProskurin A. V.^a, Post-Graduate Student, Proskurin.AV.WOF@gmail.comFavorskaya M. N.^a, Dr. Sc., Tech., Professor, favorskaya@sibsau.ru^aSiberian State Aerospace University named after academician M. F. Reshetnev, 31, Krasnoyarsky Rabochy St., 660037, Krasnoyarsk, Russian Federation

Purpose: The problem of automatic image annotation is not trivial. The training images often contain unbalanced and incomplete annotations, leading to a semantic gap between the visual features and textual description of an image. The existing methods include computationally complex algorithms which optimize the visual features and annotate a new image using all the training images and keywords, potentially reducing the accuracy. A compact visual descriptor should be developed, along with a method for choosing a group of the most informative training images for each test image. **Results:** A methodology for automatic image annotation is formulated, based on searching for a posteriori probability keyword association with a visual image descriptor. Six global descriptors combined in a single descriptor were obtained. The size of this single descriptor was reduced down to several hundred elements using principal component analysis. The experimental results showed an improvement of the annotation precision by 7% and a recall by 1%. **Practical relevance:** The compact handle visual method and automatic annotation of images based on the formation of homogeneous textual-visual groups can be used in Internet retrieval systems to improve the image search quality.

Keywords — Automatic Image Annotation, Global Visual Descriptor, Textual-Visual Groups.

References

1. Makadia A., Pavlovic V., Kumar S. A New Baseline for Image Annotation. *Proc. 10th European Conf. on Computer Vision*, Marseille, France, 2008, vol. 5304, pp. 316–329.
2. Guillaumin M., Mensink T., Verbeek J., Schmid C. TagProp: Discriminative Metric Learning in Nearest Neighbor Models for Image Auto-Annotation. *Proc. IEEE 12th Intern. Conf. on Computer Vision*, Kyoto, Japan, 2009, pp. 309–316.
3. Verma Y., Jawahar C. V. Image Annotation Using Metric Learning in Semantic Neighbourhoods. *Proc. 12th European Conf. on Computer Vision*, Florence, Italy, 2012, vol. 7574, pp. 836–849.
4. Chen M., Zheng A., Weinberger K. Q. Fast Image Tagging. *Proc. 30th Intern. Conf. on Machine Learning*, Atlanta, USA, 2013, pp. 1274–1282.
5. Blondel V. D., Guillaume J. L., Lambiotte R., Lefebvre E. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, vol. 2008, P10008.
6. Shen F., Ogura T., Hasegawa O. An Enhanced Self-Organizing Incremental Neural Network for Online Unsupervised Learning. *Neural Networks*, 2007, vol. 20(8), pp. 893–903.
7. IAPR TC-12 Benchmark. Available at: <http://www-i6.informatik.rwth-aachen.de/imageclef/resources/iaprtc12.tgz> (accessed 22 February 2016).
8. Lowe D. G. Distinctive Image Features from Scale-Invariant Keypoints. *Intern. Journal of Computer Vision*, 2004, vol. 60(2), pp. 91–110.
9. Bay H., Ess A., Tuytelaars T., Gool L. V. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 2008, vol. 110(3), pp. 346–359.
10. Yang J., Yu K., Gong Y., Huang T. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, USA, 2009, pp. 1794–1801.
11. Wang J., Yang J., Yu K., Lv F., Huang T., Gong Y. Locality-Constrained Linear Coding for Image Classification. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, USA, 2010, pp. 3360–3367.
12. Jegou H., Douze M., Schmid C., Perez P. Aggregating Local Descriptors into a Compact Image Representation. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, USA, 2010, pp. 3304–3311.
13. Proskurin A. V. Fast Local Descriptor for Scene Image Categorization. *Materialy XIX Mezhdunarodnoi nauchno-prakticheskoi konferentsii "Reshetnevskie chteniia"* [Proc. XIX Intern. Scientific Conf. "Reshetnev Readings"]. Krasnoyarsk, 2015, vol. 2, pp. 243–245 (In Russian).
14. Alcantarilla P. F., Bergasa L. M., Davison A. J. Gauge-SURF Descriptors. *Image and Vision Computing*, 2013, vol. 31(1), pp. 103–116.
15. Favorskaya M., Proskurin A. Image Categorization Using Color G-SURF Invariant to Light Intensity. *Procedia Computer Science*, 2015, vol. 60, pp. 681–690.
16. Aivazian S. A., Bukhshtaber V. M., Eniukov I. S., Meshalkin L. D. *Prikladnaya statistika. Klassifikatsiya i snizhenie razmernosti* [Applied Statistics. Classification and Reduction of Dimension]. Moscow, Finansy i statistika Publ., 1989. 607 p. (In Russian).
17. Feng S., Manmatha R., Lavrenko V. Multiple Bernoulli Relevance Models for Image and Video Annotation. *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Washington, USA, 2004, vol. 2, pp. 1002–1009.
18. Zhang S., Huang J., Li H., Metaxas D. N. Automatic Image Annotation and Retrieval Using Group Sparsity. *IEEE Transactions on Systems, Man, and Cybernetics*. Part B: Cybernetics, 2012, vol. 42(3), pp. 838–849.

ИССЛЕДОВАНИЕ КРИТИЧЕСКОГО ЗНАЧЕНИЯ ЭКРАННОГО ПАРАЛЛАКСА

Н. Н. Красильников^а, доктор техн. наук, профессор

О. И. Красильникова^а, канд. техн. наук, доцент

^аСанкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ

Цель: исследование зависимости критического значения экранного параллакса от условий наблюдения стереоскопических изображений для снижения негативных явлений, сопровождающих просмотр 3D-контента, и для выработки научно обоснованных рекомендаций по его производству. **Результаты:** предложена гипотеза, объясняющая работу зрительного механизма, который обеспечивает содружественность саккадических движений оптических осей глаз при наблюдении реальных объектов и изображений и регламентирует максимальное значение экранного параллакса. Из гипотезы следует, что этап коррекции оптической оси ведомого глаза, возникающий у части зрителей при просмотре стереоскопических изображений, приводит к осложнению процесса содружественности саккадических движений глаз и является причиной дискомфорта. Показано, что критическое значение экранного параллакса, при котором возникает эффект двоения стереоскопических изображений, зависит от их структуры, размера, а также расстояния наблюдения и может изменяться в широких пределах. При расстояниях наблюдения, превышающих 5 м, критические значения экранного параллакса изменяются прямо пропорционально расстоянию между зрителем и экраном. При создании 3D-фильмов следует ориентироваться на критические значения угловых экранного параллакса не более 17 мин. Исходя из того, что в 3D-кинематографе нет ограничений на воспроизведение достаточно больших интервалов глубины, поскольку фильм смотрится со сравнительно больших расстояний, а в 3D-телевидении такие ограничения имеются из-за малых расстояний наблюдения, для решения проблемы предлагается в качестве отображающего устройства в телевидении использовать виртуальные очки на основе смартфона. **Практическая значимость:** полученные результаты позволяют при создании 3D-контента выбирать значения экранного параллакса, гарантирующие отсутствие дискомфорта, эффекта двоения и других нежелательных артефактов. Технологии демонстрации стереоскопических изображений могут быть использованы для релаксации людей, пребывающих в замкнутых пространствах, например экипажей космических кораблей, подводных лодок и т. п., где клаустрофобия является одним из факторов, ограничивающих допустимое время работы.

Ключевые слова — стереоскопическое изображение, критическое значение экранного параллакса, содружественность саккадических движений глаз.

Введение

Технологии регистрации и отображения изображений реальных сцен, пройдя в своем развитии этапы черно-белой фотографии, черно-белого, а затем цветного кино, привели к разработке и внедрению систем объемного (3D) кино и телевидения. В этих системах в настоящее время для демонстрации 3D-изображений используется метод стереоскопии, поскольку он наиболее просто реализуется.

В настоящее время этот метод недостаточно исследован, в результате чего при его применении возникают известные проблемы, что сдерживает дальнейшее развитие 3D-кино и телевидения. Проявляется это в конечном итоге в недостаточном количестве 3D-контента, а также в негативных явлениях, сопровождающих просмотр 3D-фильмов зрителями. Некоторые зрители жалуются на ощущение дискомфорта, усталость глаз, головокружение и даже головные боли, возникающие в процессе просмотра ими 3D-кинокартин [1].

Исследования показали, что дискомфорт и другие негативные ощущения вызваны чрезмер-

но большими экранными параллаксами, оптимальная величина которых зависит от размера экрана и от расстояния между экраном и зрителем.

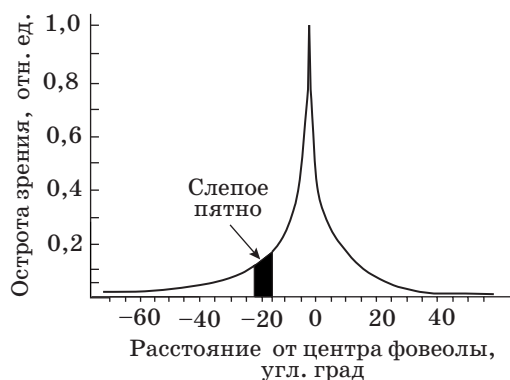
К сожалению, отсутствие теоретической базы вынуждает оператора при съемке фильма каждый раз выбирать эти параметры, ориентируясь только на имеющийся у него опыт [1]. Что же касается дальнейшего развития 3D-телевидения, то причина, сдерживающая продажи, а следовательно, и производство 3D-телевизоров, заключается в недостаточно хорошем воспроизведении глубины пространства на экранах 3D-телевизоров. Причем последнее относится не только к передачам, которые конвертируются в самом телевизоре из формата 2D в формат 3D, но и к передачам, изначально снятым в 3D-формате, т. е. в стереоскопическом формате. Целью публикуемой работы является исследование особенностей наблюдения стереоскопических изображений, учет которых позволяет уменьшить сопровождающие их негативные явления и выработать научно обоснованные рекомендации по производству 3D-контента.

Гипотеза о содружественности движения глаз при наблюдении стереоскопических изображений

Переходя к рассмотрению процессов, протекающих в зрительной системе человека при наблюдении стереоскопических изображений, начнем с того, что разрешающая способность сетчаток глаз (острота зрения) неравномерно распределена по их поверхности (рис. 1). В центре сетчатки каждого глаза имеется зона, называемая фoveолой, в которой фоторецепторы расположены с высокой плотностью и поэтому обеспечивают высокое разрешение, в то время как остальная часть сетчатки имеет низкое разрешение. Хотя угловой размер фoveолы невелик, на нее приходится около 80 % всех рецепторов сетчатки. Вследствие этого зрительной системой наблюдаемое изображение (или сцена) рассматривается последовательно по частям.

При наблюдении сцены или ее изображения оптические оси глаз совершают содружественные саккадические движения с интервалом около 50–60 мс, фиксируясь на световых границах, создаваемых объектами, составляющими наблюдаемую сцену [2, 3]. При этом происходит сканирование изображения зонами ясного видения, размер которых определяется фoveолами левого и правого глаза. В результате описанного процесса в зрительной коре формируется нейронное изображение для дальнейшего анализа. Содружественность движений глаз (а как известно, глаза человека функционально различаются между собой, поэтому выделяют ведущий и ведомый глаз) имеет ключевое значение для понимания процессов, протекающих в зрительной системе при наблюдении как реальных объектов, так и их изображений, особенно стереоскопических.

Для объяснения механизма, посредством которого обеспечивается содружественность сакка-

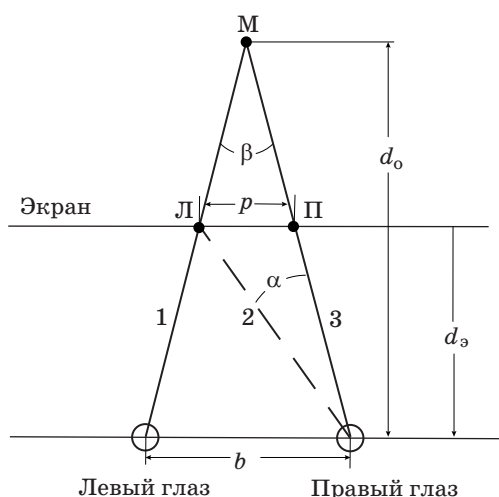


■ Рис. 1. Неравномерности остроты зрения (разрешающей способности) по поверхности сетчаток глаз

дических движений глаз, нами была выдвинута следующая гипотеза. Согласно этой гипотезе при рассматривании объекта или его 2D-изображения очередную саккаду первым начинает ведущий глаз и, заканчивая ее, фиксируется на световой границе. Далее информация о направлении оптической оси ведущего глаза и расстоянии до световой границы, которое приблизительно оценивается по его фокусировке, передается ведомому глазу, и он переводит свою оптическую ось (нацеливается) приблизительно в ту же точку пространства, что и ведущий глаз. После этого происходит корректировка направления оптической оси и фокусировка ведомого глаза таким образом, чтобы световая граница проецировалась в центр фoveолы, но так как эта корректировка невелика, на нее затрачивается мало усилий и времени, и она происходит незаметно для зрителя. Результатом описанной последовательности действий является содружественность движения глаз — одни и те же световые границы проецируются в центры фoveол обоих глаз, что принципиально важно.

Иначе обстоит дело при наблюдении стереоскопических изображений, поскольку левому и правому глазу при наблюдении предъявляются разные компоненты стереоскопического изображения, на которых положение одной и той же световой границы различается по горизонтали на величину экранного параллакса p . В этом случае, так же, как и при наблюдении 2D-изображения, ведущий глаз (пусть в рассматриваемом примере это будет левый глаз), совершая саккадическое движение, фиксируется на световой границе своего компонента стереоскопического изображения объекта таким образом, чтобы она оказалась в центре его зоны ясного видения, т. е. в центре фoveолы. Сказанное поясняет рис. 2, на котором оптическая ось ведущего глаза 1 направлена в точку L , через которую проходит граница объекта на левом компоненте стереоскопического изображения. Понятно, что эта точка расположена в плоскости экрана. Затем, в соответствии с предложенной гипотезой, ведомому глазу, как и в ранее рассмотренном случае, передается информация о пространственном положении точки L , в которую направлен ведущий глаз, и оптическая ось ведомого (в рассматриваемом примере правого) глаза 2 направляется приблизительно в эту же точку пространства.

Но информация, полученная ведомым глазом, оказывается для него ложной, поскольку ведомому глазу предъявляется совсем другой компонент стереоскопического изображения, на котором в точке L световая граница отсутствует. Напомним, что на компоненте, который предъявляется ведомому правому глазу, световая граница смещена в горизонтальном направлении



■ **Рис. 2.** Схема наблюдения стереоскопического изображения: точки *Л* и *П* — места на экране, через которые проходят вертикальные световые границы левого и правого компонентов стереоскопического изображения; точка *М* — место расположения мнимого изображения объекта; *b* — расстояние между зрачками глаз (база); *d_э* — расстояние до экрана; *d₀* — расстояние до объекта; β — угловой параллакс; α — угловое значение экранного параллакса

на величину экранного параллакса p относительно световой границы компонента, предъявляемого ведущему глазу, и находится в точке *П*. Таким образом, для ведомого глаза световая граница будет проецироваться не в центр фовеолы, а в область, смещенную относительно центра на величину, которая определяется величиной экранного параллакса. В этом случае, как и ранее, ведомый глаз также корректирует положение своей оптической оси в пространстве так, чтобы проекция световой границы оказалась в центре его фовеолы, для чего оптическая ось ведомого глаза смещается на величину значения углового экранного параллакса α , оказывается в положении, показанном на рис. 2 линией 3, и происходит «захват» световой границы. Однако для этого затрачивается больше усилий и времени, чем в предыдущем случае, когда изображение не стереоскопическое, и тем больше, чем больше экранный параллакс. Вследствие этого у зрителя возникает ощущение дискомфорта, возрастающее с увеличением экранного параллакса.

При захвате световой границы продолжения оптических осей глаз сходятся в точке *М* пересечения осей зрения, соответствующей расположению световой границы мнимого изображения объекта (см. рис. 2). При этом угловой параллакс β передается в зрительную кору, и по нему определяется расстояние до световой границы этого мнимого изображения объекта. В случае, показанном на рис. 2, объект кажется расположен-

ным дальше, чем плоскость экрана, при этом стереоскопическое изображение воспринимается зрителем как объемное.

Если значения экранного параллакса превышают некоторое критическое значение $p_{кр}$, требуемая корректировка оказывается чрезмерно большой, и захвата световой границы не происходит; в результате этого возникает эффект так называемого двоения, стереоскопическое изображение распадается на два взаимно смещенных, наложенных друг на друга плоских изображения.

Из изложенного следуют два вывода.

Во-первых, поскольку зоны ясного видения не имеют четко выраженных границ, то критические значения экранных параллаксов $p_{кр}$, при которых возникает эффект двоения, также будут иметь существенный разброс.

Во-вторых, при расстояниях наблюдения, превышающих 5 м, когда глубина резкости уже практически не зависит от расстояния до экрана, критические значения экранных параллаксов будут прямо пропорциональны расстоянию между зрителем и экраном $d_э$, а их угловые значения практически не будут зависеть от расстояния до экрана.

Экспериментальное исследование зависимости критического значения экранного параллакса от условий наблюдения стереоскопического изображения

Выдвинутая нами гипотеза, предназначенная для объяснения механизма формирования координаты глубины световой границы мнимого изображения объекта при рассматривании стереоскопического изображения, была проверена экспериментально. Проверка включала в себя три серии экспериментов.

В *первой серии* экспериментов исследования выполнялись с использованием специальных тестовых стереоскопических изображений, на которых были представлены отрезки линий, различным образом ориентированные в пространстве. Цель экспериментов заключалась в определении зависимости критического значения экранного параллакса $P_{кр}$ от толщины линий и расстояния от наблюдателя до экрана $d_э$.

В этих исследованиях в качестве отображающего устройства использовался 3D-телевизор Samsung VE40F6650AB и светоклапанные очки, так как в этом типе телевизоров применен эклипсный метод разделения ракурсов. На экран телевизора в формате МРО выводились стереоскопические тестовые изображения отрезков линий толщиной 1, 2, 4, 8 и 14 пикселей, верхние части которых были расположены в плоскости экрана, а нижние — либо перед экраном, либо за ним в зависимости от используемого тестового

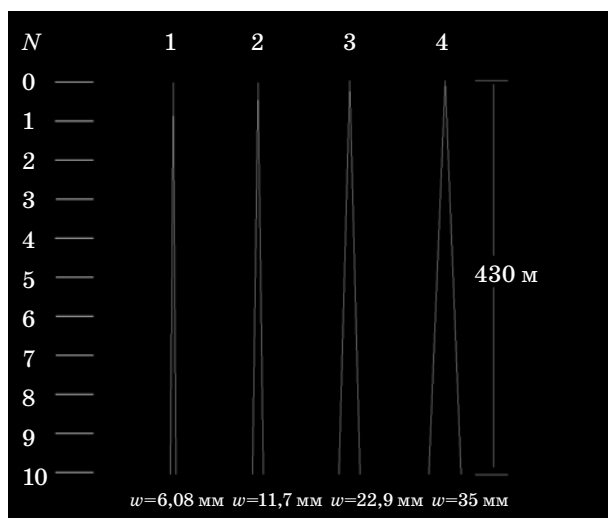
изображения. Отрезки линий были расположены в вертикальных плоскостях, перпендикулярных плоскости экрана телевизора. Как выглядит на экране телевизора одно из использованных в эксперименте тестовых изображений с отрезками линий толщиной 1 пиксель, если смотреть на него непосредственно без светоклапанных очков, показано на рис. 3. Обращаем внимание на то, что благодаря отсутствию светоклапанных очков на экране одновременно видны оба компонента стереоскопического изображения, предназначенные для наблюдения и левым, и правым глазом, в виде расходящихся вследствие экранного параллакса отрезков прямых линий.

В процессе экспериментальных исследований зрителю, который смотрел на экран через светоклапанные очки с заданного расстояния d_3 , предъявлялось одно из тестовых изображений. Зритель по шкале, имеющейся на левой стороне тестового изображения, определял, на каком уровне (при каком значении N) исчезает стереоскопический эффект и возникает эффект двоения линии. Величина критического значения экранного параллакса $p_{кр}$, при котором происходило раздвоение наблюдаемой линии, вычислялась по формуле

$$p_{кр} = wN/10.$$

В экспериментах приняли участие 11 чел. Всего было выполнено более 1500 измерений.

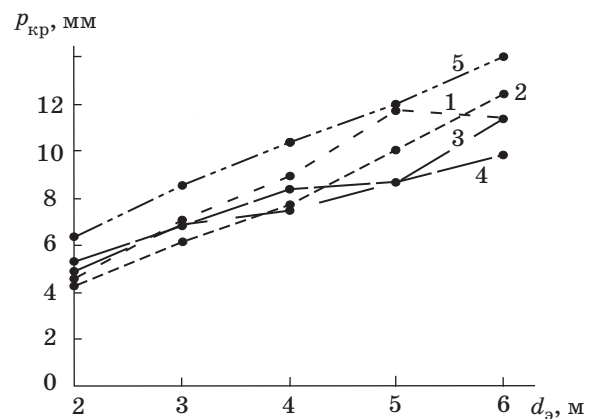
В процессе выполнения экспериментов все испытуемые отмечали трудность фиксации и некоторую неопределенность уровня N , при котором возникает эффект «двоения». Этот эффект то появ-



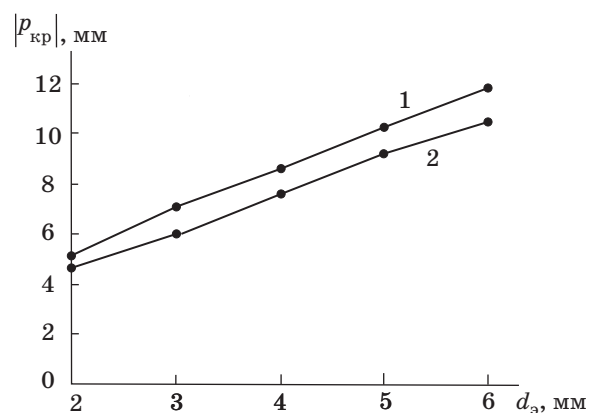
■ Рис. 3. Пример тестового стереоскопического изображения, если смотреть на него без светоклапанных очков, с линиями толщиной 1 пиксель: w — расстояние между линиями

лялся, то исчезал. В итоге даже при усреднении по всем испытуемым результатов измерения $p_{кр}$ получился значительный разброс данных, о величине которого можно судить по графикам (рис. 4), построенным для случая, когда нижние части отрезков линий теста были расположены за экраном. Графики убеждают, что какая-либо явно выраженная зависимость $p_{кр}$ от толщины отрезков линий теста отсутствует, поэтому наблюдаемый разброс следует отнести к неопределенности установления уровня N , на которую ссылались испытуемые.

На рис. 5 представлены графики зависимости абсолютных величин критических значений экранного параллакса от расстояния наблюдения, усредненные по всем испытуемым, принявшим участие в эксперименте, отдельно для тестовых



■ Рис. 4. Зависимости 1–5 критического значения экранного параллакса $p_{кр}$ от расстояния наблюдения d_3 для линий тестового изображения толщиной 1, 2, 4, 8 и 14 пикселей соответственно, усредненные по всем наблюдателям



■ Рис. 5. Зависимости абсолютных величин критических значений экранного параллакса $p_{кр}$ от расстояния наблюдения d_3 : 1 — отклонение нижней части тестовой линии от зрителя; 2 — отклонение нижней части тестовой линии в сторону зрителя

вых изображений, в которых концы отрезков линий были расположены перед и за экраном.

По данным, приведенным на рис. 5, для различных расстояний наблюдения были рассчитаны угловые размеры критических значений экранных параллаксов по формуле

$$\alpha_{кр} = \arctg(p_{кр}/d_3),$$

которые в виде зависимостей от расстояния до экрана представлены на рис. 6.

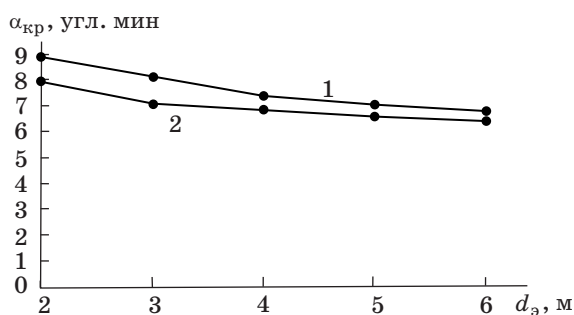
Обратим внимание на то, что измеренные угловые размеры критических значений экранных параллаксов в описанных экспериментах оказались существенно меньше диаметра фовеолы, величина которой по различным литературным источникам находится в пределах 35–50 мин. Кроме того, как и было предсказано предложенной гипотезой, при расстояниях наблюдения, превышающих 5 м, критические значения экранных параллаксов изменяются прямо пропорционально расстоянию между зрителем и экраном d_3 , а их угловые значения практически от него не зависят.

Во второй серии экспериментов исследования выполнялись с использованием сюжетных стереоскопических изображений, существенной особенностью которых, в отличие от тестовых изображений с отрезками линий, являлся элемент неопределенности в положении световых границ, создаваемых изображениями реальных объектов на сложном неравномерном фоне. Цель экспериментов заключалась в определении зависимости критического значения экранного параллакса $p_{кр}$ от расстояния наблюдения d_3 .

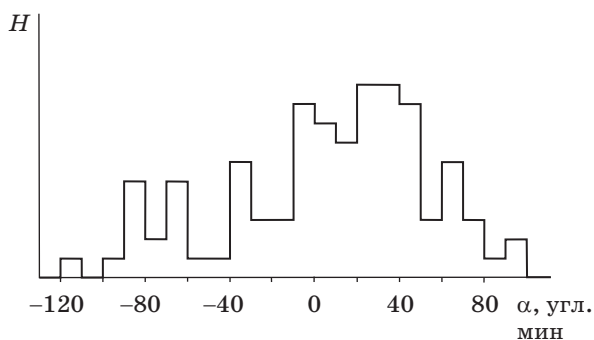
Для этих исследований было специально разработано приложение, создававшее из исходного 2D-изображения левый и правый компоненты стереопары путем простого смещения красного и голубого компонентов исходного изображения влево и вправо относительно друг друга на задан-

ное количество пикселей. При наблюдении созданного таким образом стереоизображения зритель с использованием анаглифных очков видел на экране монитора компьютера по-прежнему исходное, плоское изображение, но смещенное по глубине. Приложение позволяло зрителю в процессе наблюдения простым изменением положения движка полосы прокрутки изменять значение параллакса β и, как следствие, изменять величину смещения по глубине наблюдаемого изображения. При выполнении эксперимента в задаче зрителя входило определение такого углового значения экранного параллакса $\alpha_{кр}$, при котором возникает эффект двоения изображения. Так же, как и в описанных выше экспериментах, зрители отмечали трудность четкого определения порога, когда возникал эффект двоения изображения, но этот порог был существенно выше, чем при экспериментах с линиями. И положительные, и отрицательные угловые критические значения экранного параллакса, полученные в результате этих измерений, во всем интервале дальностей наблюдения приближались к одному градусу, а в некоторых случаях и превышали его. Однако еще задолго до того, как возникал эффект двоения изображения, у зрителя появлялось ощущение дискомфорта. Зрители отмечали, что эффект двоения даже после возникновения присутствовал не постоянно, а то возникал, то исчезал. Объяснить это можно тем, что при наблюдении сюжетных изображений задача зрителя усложняется неопределенностью положения световых границ в поле его зрения. Интересно отметить, что в статье [4] со ссылкой на работу [5] приводится угловое значение критического параллакса $\alpha_{кр}$, равное 60–70 мин. Тем не менее мы полагаем, что при создании 3D-фильмов следует ориентироваться на угловые критические значения экранных параллаксов не более 17 мин, поскольку при больших значениях возникает ощущение дискомфорта и, как следствие, зрители быстрее утомляются.

Третья серия экспериментов заключалась в измерении максимальных значений угловых параллаксов, которые имеют место в реальных 3D-фильмах. Для выполнения экспериментов были выбраны четыре 3D-фильма: «Сталинград», «Титаник», «Хранитель времени» и «Шрек», — которые различались как технологией изготовления, так и жанром. Фильмы «Сталинград» и «Хранитель времени» были сняты в версии 3D стереокамерой; версия 3D фильма «Титаник» была получена путем конвертации из 2D-формата, особенностью фильма «Шрек» является то, что он анимационный [6–8]. Из этих фильмов были отобраны 60 кадров для измерения в каждом из них одного максимального положительного и одного отрицательного параллакса, которое



■ Рис. 6. Зависимости угловых критических значений экранных параллаксов $\alpha_{кр}$ от расстояния наблюдения d_3 : 1 — отклонение нижней части тестовой линии от зрителя; 2 — отклонение нижней части тестовой линии в сторону зрителя



■ Рис. 7. Гистограмма H распределения угловых значений экранных параллаксов α

проводилось с помощью специального приложения, разработанного для этой цели. Измерения параллаксов заключались в совмещении световых границ левого и правого компонентов изображений посредством регулировки движка полосы прокрутки. Полученные при измерениях значения параллаксов в пикселях записывались, а затем использовались для расчета угловых значений параллаксов. Угловые значения параллаксов зависят также от размера изображения и расстояния, с которого производится наблюдение. Эти данные мы взяли для кинотеатра «Аврора», полагая их типичными для 3D-кинотеатров: размер экрана 14×7 м, расстояние наблюдения из среднего ряда $\alpha_0 = 6,45$ м. По результатам измерений была рассчитана гистограмма распределения угловых значений экранных параллаксов α (рис. 7).

В результате проведенных исследований было определено, что в этих фильмах в основном преобладали положительные параллаксы, составляющие $2/3$ от всех просмотренных. Кроме того, среди измеренных значений максимальное значение положительного параллакса составляло 105 мин, а отрицательного — 202 мин, хотя в основном преобладали, как это видно из гистограммы, значения меньше одного градуса. В связи с этим уместно сослаться на статью [1], в которой рассматриваются проблемы, связанные с отрицательными артефактами при просмотре 3D-фильмов, и в частности говорится, что комфортным для кинотеатров считается среднее значение параллакса порядка $0,5\%$ от ширины экрана. Пересчет приведенной величины в угловое значение экранного параллакса для кинотеатра «Аврора» при просмотре из среднего ряда дает значение $\alpha_{кр} = 15$ мин.

Заключение

Предложенная нами гипотеза, объясняющая механизм содружественного саккадического движения оптических осей глаз, не только по-

зволяет просто рассчитать критическое значение экранного параллакса, гарантирующее отсутствие дискомфорта и тем более двоения изображения, исходя из рекомендуемого углового размера критического значения экранного параллакса и расстояния наблюдения стереоскопического изображения. Эта гипотеза с большой долей уверенности указывает на причину дискомфорта, возникающего у части зрителей при просмотре стереоскопических изображений с большим значением экранного параллакса, которая заключается в том, что процесс содружественности саккадических движений глаз в данном случае осложнен этапом коррекции оптической оси ведомого глаза.

Как известно, в 3D-кинематографе не возникает ограничений на воспроизведение достаточно больших интервалов глубины благодаря тому, что стереоскопический фильм смотрится со сравнительно больших расстояний, в то время как в 3D-телевидении вследствие того, что наблюдение стереоскопических изображений осуществляется с небольших расстояний (часто с двух-трех метров), воспроизведение больших интервалов глубины невозможно [1]. Выходом из ситуации может служить индивидуальный просмотр телевизионных передач не на экранах 3D-телевизоров, а с использованием сравнительно недорогих приборов, аналогичных тем, которые в настоящее время широко рекламируются, так называемых 3D-очков виртуальной реальности [9]. Применение таких приборов для индивидуального просмотра 3D-телевизионных передач позволяет также решить еще одну проблему — проблему дискомфорта, головокружения и даже головной боли, возникающую у некоторых людей при просмотре 3D-фильмов, путем индивидуальной настройки параметров демонстрируемого изображения.

Отметим, что технологии получения и демонстрации стереоскопических изображений помимо традиционного применения в кино, телевидении и всевозможных аттракционах могут быть эффективно использованы для релаксации людей, долго находящихся в замкнутом пространстве, например на борту космического корабля, подводной лодки и т. п., где из-за клаустрофобии ограничивается допустимое время работы. В отличие от ставшего уже традиционным метода создания виртуальных миров, в котором применяются 3D-изображения, в рассматриваемом случае используются стереоскопические изображения. Данный метод позволяет при сравнительно небольших вычислительных затратах получать высококачественные объемные изображения представляемой сцены с возможностью ее кругового обзора и ограниченного перемещения.

Литература

1. Почему от 3D болит голова. Ч. 4: Параллакс. <https://geektimes.ru/post/248970/> (дата обращения: 22.03.2016).
2. Физиология сенсорных систем. Ч. I: Физиология зрения / под ред. Г. В. Гершуни. — Л.: Наука, 1971. — 416 с.
3. Ярбус А. Л. Роль движения глаз в процессе зрения. — М.: Наука, 1965. — 166 с.
4. Федоров С. Природа объема // Журнал «625». 2010. № 3. С. 20–25.
5. Стереотелевидение (черно-белое и цветное): учеб. пособие для вузов связи / под ред. П. В. Шмакова. — М.: Связь, 1968. — 207 с.
6. Конвертация 2d в 3d. Что такое псевдостерео? <http://www.televizor-3d.ru/konvertaciya-2-d-v-3d.html> (дата обращения: 22.03.2016).
7. Почему «Титаник» стоит смотреть в стерео 3D. <http://total3d.ru/diy/92707/> (дата обращения: 22.03.2016).
8. Красильников Н. Н., Красильникова О. И. Метод конвертации 2D-изображения в стереоскопическое 3D-изображение // Оптический журнал. 2014. Т. 82. № 2. С. 20–28.
9. Шлемы и очки виртуальной реальности для ПК. <http://oculus-rift.ru/shlemy-i-ochki-virtualnoj-realnosti/> (дата обращения: 22.03.2016).

UDC 612.843,721

doi:10.15217/issn1684-8853.2016.2.19

Study of Critical Values of On-Screen Parallax

Krasilnikov N. N.^a, Dr. Sc., Tech., Professor, NNKrasilnikov@yandex.ruKrasilnikova O. I.^a, PhD, Tech., Associate Professor, OIKrasilnikova@yandex.ru^aSaint-Petersburg State University of Aerospace Instrumentation 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

Purpose: Studying how the critical value of on-screen parallax depends on the observation conditions for stereoscopic images is important to reduce the negative phenomena accompanying the viewing of 3D content and to develop evidence-based recommendations for its production. **Results:** We have proposed a hypothesis which explains the functioning of the visual mechanism providing mutual saccadic movements of the optical axes of human eyes when observing real objects and images. This mechanism regulates the maximum value of the screen parallax. From the hypothesis it follows that the complication of mutual saccadic eye movement at the stage of correcting the optical axis of the slave eye which some people experience when viewing stereoscopic images is the cause of their discomfort. It is shown that the critical value of screen parallax at which the effect of ghosting in stereoscopic images occurs depends on the image structure, size and observation distance, and can vary within wide limits. When the observation distance is greater than 5 m, the critical values of screen parallaxes are directly proportional to the distance between the viewer and the screen. Developers of 3D films should be oriented to the critical values of angle screen parallaxes no more than 17 minutes. Based on the fact that 3D cinema has no restrictions on reproducing sufficiently large depth intervals (because a film is watched at a relatively large distance, and in 3D TV such restrictions take place due to small observation distances), to solve the problem in television we propose using virtual glasses on the base of a smartphone as a display device. **Practical relevance:** The obtained results allow us, when creating 3D content, to choose the values of screen parallax, ensuring the absence of discomfort, ghosting effect and other unwanted artifacts. Technologies of demonstrating stereoscopic images can be used for the relaxation of people staying in confined spaces, such as spaceships, submarines, etc., where claustrophobia is one of the factors limiting the allowable time.

Keywords — Stereoscopic Image, Critical Value of Screen Parallax, Mutual Saccadic Eye Movement.

References

1. *Pochemu ot 3D bolit golova. Ch. 4. Parallaks* [Why Headache from 3D. Part 4. Parallax]. Available at: <http://savepearlharbor.com/?p=255241> (accessed 22 March 2016).
2. *Fiziologiya sensornykh sistem. Ch. I. Fiziologiya zreniya* [Physiology of Sensory Systems. Part One. Vision Physiology]. Ed. by G. B. Gershuni. Leningrad, Nauka Publ., 1971. 416 p. (In Russian).
3. Yarbus A. L. *Rol' dvizheniya glaz v protsesse zreniya* [The Role of Eye Movements in the Visual Process]. Moscow, Nauka Publ., 1965. 166 p. (In Russian).
4. Fyodorov S. The Nature of Volume. *Zhurnal "625"*, 2010, no. 3, pp. 20–25 (In Russian).
5. *Stereotelevidenie (cherno-beloe i tsvetnoe)*. Ed. by P. V. Shmakova. Moscow, Sviaz' Publ., 1968. 207 p. (In Russian).
6. *Konvertatsiya 2d v 3d. Chto takoe psevdostereo?* [Converting 2D to 3D. What is Pseudostereo?]. Available at: <http://www.televizor-3d.ru/konvertaciya-2-d-v-3d.html> (accessed 22 March 2016).
7. *Pochemu "Titanik" stoit smotret' v stereo 3D* [Why "Titanic" is Worth Watching in Stereo 3D]. Available at: <http://total3d.ru/diy/92707/> (accessed 22 March 2016).
8. Krasilnikov N. N., Krasilnikova O. I. The Method of Converting a 2D Image into a Stereoscopic 3D Image. *Opticheskii zhurnal*, 2014, vol. 82, no. 2, pp. 20–28 (In Russian).
9. *Shlemy i ochki virtual'noi real'nosti dlia PK* [Helmets and Glasses of Virtual Reality for PC]. Available at: <http://oculus-rift.ru/shlemy-i-ochki-virtualnoj-realnosti/> (accessed 22 March 2016).

МЕТОДЫ И СРЕДСТВА АНАЛИЗА НАДЕЖНОСТИ СТРУКТУРНЫХ БЛОКОВ С РЕЗЕРВИРОВАНИЕМ И ПЕРИОДИЧЕСКИМ ВОССТАНОВЛЕНИЕМ ИНФОРМАЦИИ НА РАЗЛИЧНЫХ ЭТАПАХ ПРОЕКТИРОВАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

И. В. Егоров^а, аспирант

В. Ф. Мелехин^а, доктор техн. наук, профессор

^аСанкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, РФ

Постановка проблемы: проектирование отказоустойчивых вычислительных систем, работающих в условиях повышенной радиации, требует оценки влияния сбоев, заключающихся в искажении информации при попадании заряженной частицы. Средства анализа надежности таких систем должны учитывать возникновение таких сбоев, а также моделировать влияние средств периодического восстановления искаженной информации, применяемых для повышения характеристик надежности. **Цель:** определение границ применимости методов анализа надежности, основанных на использовании аналитических и имитационных моделей, для оценки надежности систем с периодическим восстановлением; сравнение результатов анализа показателей надежности восстанавливаемой системы, полученных с помощью аналитической и имитационной моделей. **Результаты:** с помощью аналитической и имитационной моделей произведен расчет и анализ времени наработки восстанавливаемой системы до отказа. Выявлено, что наличие детерминированных процессов восстановления ограничивает применимость традиционных методов оценки надежности, таких как построение марковской модели и логико-вероятностные методы. Определена необходимость учета влияния программного обеспечения при проектировании систем с периодическим восстановлением. **Практическая значимость:** благодаря проведенным исследованиям определены области применения моделей надежности разного вида на различных этапах проектирования отказоустойчивых восстанавливаемых вычислительных систем.

Ключевые слова — имитационное моделирование, надежность восстанавливаемых систем, радиационно-стойкая вычислительная система, марковская модель, структурное резервирование, информационная избыточность.

Введение

Из работ, посвященных исследованию влияния радиации на полупроводниковые структуры [1–3], известно, что попадание заряженных частиц наиболее часто приводит к появлению ложных импульсов на выходах логических элементов и сохранению искаженной информации в триггерах, иными словами, к возникновению «мягких» (информационных) отказов. При этом аппаратура продолжает находиться в работоспособном состоянии, а это значит, что за счет имеющейся временной избыточности и специальной организации вычислительного процесса можно обеспечить периодическое восстановление поврежденной информации и вернуть уровень работоспособности устройства к начальному состоянию.

Статистические исследования [4] показывают, что полное разрушение структуры полупроводника, которое влечет окончательный выход элементов вычислительной системы (ВС) из строя, происходит в сотни раз реже «мягких» отказов. Это меняет взгляд на организацию проектирования отказоустойчивых систем, поскольку распространенные в традиционных подходах методы структурного резервирования, используемые

при борьбе с невосстанавливаемыми отказами, должны быть дополнены средствами периодического восстановления поврежденной информации, что существенно повысит надежность системы. Такой подход приводит к новым архитектурным решениям.

В работах [5, 6] предложен подход к проектированию ВС, основанный на представлении системы в виде сети функциональных блоков (конечного автомата с памятью или модуля памяти), внутри которых реализуются средства защиты от «мягких» отказов путем периодического восстановления информации с помощью структурного резервирования для конечных автоматов и использования корректирующих кодов с исправлением ошибок для блоков памяти.

Организация восстановления влечет за собой новые требования к анализу надежности и к моделям надежности. Аналитические модели, основанные на вероятностной оценке возникновения отказов, применимы для случаев, когда среднее время между двумя «мягкими» отказами значительно меньше периода восстановления информации. При недостаточной временной избыточности в системе эти ограничения могут не выполняться. В такой ситуации требуются модели надежности, учитывающие регулярные процессы

восстановления и влияние логики программного обеспечения на процессы в системе. Этим обосновано применение метода оценки, базирующегося на программной имитации процессов, протекающих в системе (таких как возникновение отказов, восстановление информации, действие вычислительного процесса и т. д.), т. е. имитационных моделей.

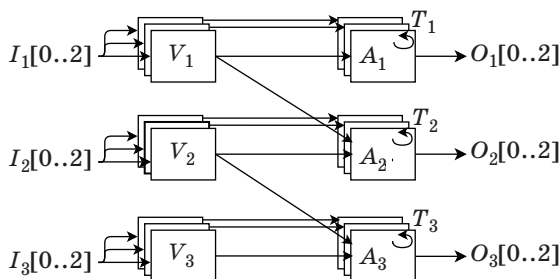
В настоящей работе предлагается реализация имитационной модели надежности, позволяющей учитывать влияние процессов восстановления на работоспособность системы. С помощью данной имитационной модели производится оценка системы со структурной избыточностью и восстановлением.

Объект исследования

Для иллюстрации использования аналитических и имитационной моделей оценим структурную надежность фрагмента структуры (рис. 1).

Представленная структура является частным случаем обобщенной структуры (см. рис. 2), используемой при проектировании отказоустойчивых ВС, основанном на разбиении системы на независимые блоки с периодическим восстановлением работоспособного состояния блока [7]. Такое построение структуры обосновано для невозстанавливаемых систем и соответствует «доменной» организации, допускающей масштабирование уровня резервирования. Структура обеспечивает блокирование распространения отказов за пределы блока как для невозстанавливаемых, так и для «мягких» отказов.

Выбор этой структуры обусловлен также тем, что именно для нее в работе [8] получена аналитическая модель расчета показателей надежности с учетом периодического восстановления «мягких» отказов. С этой аналитической моде-



■ **Рис. 1.** Исследуемый фрагмент структуры: I_1, I_2, I_3 — входные данные; V_1, V_2, V_3 — группы троированных мажоритаров; A_1, A_2, A_3 — группы троированных блоков типа «конечный автомат с памятью» с периодическим восстановлением (каждая тройка блоков имеет свой собственный период восстановления T_1, T_2, T_3); O_1, O_2, O_3 — выходы троированных блоков

лью будем сравнивать расчеты на имитационной модели. Вопрос о целесообразности модификации структуры применительно к случаю прева-лирования мягких отказов требует отдельного рассмотрения.

Если в каком-либо блоке произошел один отказ за период восстановления, он никак не скажется на функционировании других, связанных с его выходами, блоков благодаря мажоритарам, расположенным на их входах.

Система считается работоспособной, если на каждом из выходов O_1, O_2, O_3 выставлены корректные данные. Это условие выполняется, если на выходе каждого блока присутствует не более одного искаженного сигнала из трех.

На примере из рис. 1 продемонстрируем расчет интенсивности потока отказов на выходах O_1, O_2, O_3 в зависимости от интенсивности восстановления информации в блоках $A_N = \lambda_{rn}$ при заданных значениях:

- интенсивность собственных отказов каждого экземпляра в тройке блоков $A_N = \lambda_n$;
- интенсивность собственных отказов каждого мажоритара в тройке $V_N = \delta_n$.

Подходы к аналитическому расчету

Аналитическая оценка данной системы может быть произведена одним из следующих методов:

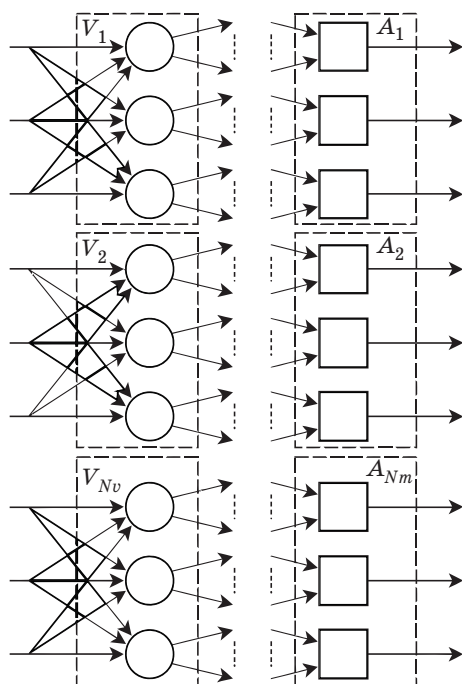
- комбинаторной оценкой возникновения различных комбинаций отказов в блоках и получением аналитической формулы для расчета общей вероятности отказа системы [8];
- построением марковской модели и ее решением [9];
- построением логической функции работоспособного состояния и ее решением логико-вероятностными методами [10, 11].

Комбинаторная оценка

В работе [8] выводится обобщенная формула, оценивающая надежность отказоустойчивой структуры (рис. 2), состоящей из троированных мажоритаров V_i на входе и подключенных к ним троированных периодически восстанавливаемых блоков A_k . Данная структура является обобщением исследуемой (см. рис. 1) с соответствующими наименованиями элементов.

Зависимость интенсивности потока отказов на выходе блока от параметров восстановления определяется по формуле

$$\lambda_s = 3 \sum_{j \in 1:N_m} \lambda_j^2 t_{Rj} + 3 \sum_{j \in 1:N_v} \delta_j^2 K_j t_{R_{\max(j)}} + 6 \sum_{j,k \in 1:N_v} \delta_j \delta_k K_{jk} t_{R_{\max(j,k)}} + 6 \sum_{i \in 1:N_v} \delta_i \lambda_i t_{Rj}. \quad (1)$$



■ Рис. 2. Обобщенная структура соединений отказоустойчивых блоков

В формуле (1) использованы следующие обозначения:

λ_s — интенсивность отказов на выходах структуры (см. рис. 2);

N_m — количество троированных блоков;

N_v — количество троированных мажоритаров;

K — коэффициент кратности периодов обновления в блоках ($1 \leq K \leq 2$). При кратных периодах обновления $K = 1$;

$\text{out}: V \rightarrow 2^M$ — функция, определяющая множество троек узлов, подключенных к выходам заданной тройки мажоритаров;

λ_j — интенсивность собственных отказов экземпляров блока j ;

δ_j — интенсивность отказов мажоритаров j . Стоит отметить, что эта величина отражает только отказы мажоритаров, оказавшие непосредственное влияние на подключенный к ним блок, т. е. повлекшие изменение состояния памяти блока. При рассмотрении «мягких» отказов необходимо учитывать, что не каждое попадание заряженной частицы в область мажоритаров влечет за собой сбой в работе подключенного автомата. Наведенные импульсы, не сохранившиеся в течение всего времени удержания запоминающего элемента, не вызовут запоминания искаженного бита. Также при оценке влияния мажоритаров необходимо учитывать особенности их структуры — как минимум тот факт, что вероятность попадания частицы прямо пропорциональна площади, занимаемой элементом на кристалле, которая для мажоритаров относительно мала.

Вышеперечисленные причины уменьшают вероятность отказа системы из-за сбоев в мажоритаровых и уменьшают их структурную значимость с точки зрения надежности;

t_{Rj} — период восстановления блока j ;

$t_{R_max(j, k)}$ — максимальный из периодов восстановления блоков j и k .

При получении формулы (1) принимались следующие допущения:

— в элементе в течение периода восстановления возникает не более двух отказов (предположение о наличии достаточного резерва времени для обеспечения нужной интенсивности регулярного восстановления);

— отсутствуют невосстанавливаемые отказы.

Один из основных выводов, полученных из формулы (1): интенсивность отказов на выходе троированных блоков находится в прямой линейной зависимости от периода восстановления блоков.

Марковская модель

При проведении комбинаторной оценки приходится ограничивать количество рассматриваемых событий, возникающих за период восстановления, чтобы сократить количество анализируемых комбинаций и упростить тем самым итоговое выражение. Этого ограничения можно избежать, если выделить все возможные состояния системы (полностью работоспособное и состояния с частичной деградацией в результате отказов экземпляров блоков и мажоритаров) и представить систему в виде марковской модели. В этом случае можно получить точный результат путем решения системы алгебраических уравнений Чепмена — Колмогорова [9].

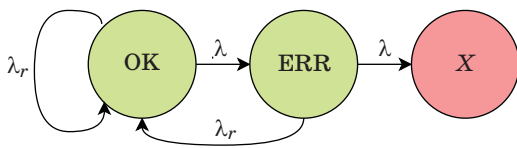
Однако марковская модель накладывает ограничения следующего рода:

— величины интервалов времени между отказами и периодов между восстановлениями подчиняются экспоненциальному закону распределения случайных величин;

— рост сложности системы ведет к резкому увеличению числа состояний, что приводит к усложнению системы уравнений и ее решения.

Первое ограничение значительно затрудняет анализ систем с детерминированным восстановлением. К примеру, рассмотрим поведение некоторого элемента системы, устойчивого к появлению однократного сбоя за детерминированный период восстановления $T_{\text{восст}} = 1/\lambda_r$. Поток отказов пуассоновский с интенсивностью λ .

Для построения марковской модели такой системы необходимо принять допущение об экспоненциальном законе распределения интервалов между событиями как потока восстановления, так и потока отказов. Модель будет иметь вид, представленный на рис. 3.



■ **Рис. 3.** Модель состояний элемента, устойчивого к однократному сбою

Пусть $\lambda_r = 1,0$, $\lambda = 0,2$. Тогда оценка среднего времени T до перехода в неработоспособное состояние X из полностью работоспособного состояния OK и состояния с частичной деградацией ERR (error) на основании марковской модели дает результат $T = 33,78$ (некорректная оценка).

При использовании комбинаторного метода (корректной оценки) необходимо оценить вероятность возникновения двух и более отказов за период восстановления 1,0. Так как поток отказов пуассоновский, эта вероятность приближенно

$$\text{равна } P_0 = \sum_{n=2}^{\infty} P(n) = \sum_{n=2}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \approx 0,018. \text{ Данная ве-}$$

личина определяет среднее количество периодов восстановления, прошедших до возникновения первого сбоя, и позволяет оценить среднее время наработки до отказа $T = \frac{T_{\text{восст}}}{P_0} = \frac{1,0}{0,018} = 55,56$.

Расхождение в результатах оценки среднего времени наработки двумя методами иллюстрирует некорректность допущения об экспоненциальном законе распределения периода восстановления, использованного при построении марковской модели (при переходе из состояния в состояние марковская модель «забывает» предысторию перехода и, следовательно, информацию о том, сколько времени осталось до наступления следующего момента восстановления).

Логико-вероятностные методы

Рассмотрение использования логико-вероятностных методов [10, 11] выходит за рамки данной статьи. Лишь отметим, что применение этих методов по сравнению с марковскими моделями зачастую избавляет от трудоемкого построения модели состояний системы, но сохраняет ограничения на закон распределения интенсивностей отказов и восстановлений, характерные для применения марковской модели, а также добавляет требование независимости событий отказов, возникающих в элементах системы.

Применение имитационной модели

Цель применения имитационной модели — получение характеристик надежности для систем, не укладывающихся в ограничения традиционных аналитических моделей.

Для этого на ЭВМ производится моделирование последовательности возникших в системе событий и ее реакций на них (эксперимент). Эксперимент заканчивается при наступлении определенного условия (например, при переходе системы в неработоспособное состояние или по истечении времени моделирования).

В качестве имитационных моделей надежности в различных работах предлагаются либо модели общего назначения (к примеру, сети Петри), либо специализированные, ориентированные на оценку надежности специфических систем [12].

Имитационная модель, предложенная в данной работе, имеет сходства с марковской:

- для ее построения задаются состояния системы, определяемые последовательностью возникших в системе сбоев (состояния частичной деградации);

- между состояниями задаются переходы, главные причины которых — возникновение отказов и событий восстановления. Вес перехода задает интенсивность возникновения данного события.

С другой стороны, имитационная модель расширяет возможности марковских моделей:

1. Интервалы времени между переходами не обязательно подчинены экспоненциальному закону распределения случайной величины. В общем случае имеется возможность выбора закона распределения и параметров функции распределения для данного закона. В частности, может быть задан детерминированный интервал времени, через который осуществляется переход, что полезно для моделирования процессов восстановления.

2. Многие события, возникающие в системе, не зависят от ее текущего состояния. Например, восстановление происходит через фиксированные моменты времени, в каком бы состоянии система не находилась. Такие явления можно задать с помощью генератора глобальных событий (Global Events Generator). Также в модели определяется реакция системы на эти события, которая в общем случае зависит от текущего состояния.

3. Марковская модель требует определения всех возможных состояний системы и переходов между ними. Для системы из рис. 1 требовалось бы выделить отдельное состояние для каждой возможной комбинации отказавших элементов — мажоритаров и блоков. Модель в этом случае содержала бы 22 состояния. При возрастании сложности системы количество ее состояний возрастало бы еще значительно. Для устранения упомянутого ограничения предлагается представить имитационную модель в виде сети независимых графов состояний, взаимосвязь между которыми обеспечивается путем передачи сигналов (именованных событиями) от одного графа

к другому. Такого рода событие порождается графом при возникновении в нем перехода из состояния в состояние. Данное событие может стать причиной перехода между состояниями в других графах. Такой подход позволяет перейти от моделирования сложной системы к моделированию нескольких простых со значительно меньшим количеством состояний.

4. Следствием из п. 3 является то, что работоспособность сети графов задается в виде логической функции (Health Function), зависящей от текущего состояния заданного подмножества графов (непосредственно определяющего работоспособность системы), входящих в состав сети.

Результатом расчета модели является:

- вероятностная функция работоспособности;
- среднее время наработки до отказа;
- оценка погрешности рассчитанных характеристик.

Пример расчета

При расчете системы (см. рис. 1) используются следующие допущения и относительные значения:

- поток отказов в блоках и мажоритарных является пуассоновским;
- отсутствуют невосстанавливаемые отказы в элементах;
- интенсивности отказов всех экземпляров блоков A_1, A_2, A_3 равны $\lambda = \lambda_1 = \lambda_2 = \lambda_3 = 1,0$;
- интенсивности отказов всех экземпляров мажоритаров V_1, V_2, V_3 равны $\delta = \delta_1 = \delta_2 = \delta_3 = 0,01$;
- период восстановления блоков A_1, A_2, A_3 детерминирован и равен $t_r = 1/\lambda_r$.

Аналитическая оценка комбинаторным методом.

Для аналитической оценки применим формулу (1), которая для данной структуры имеет вид

$$\begin{aligned} \lambda_s = & 3(\lambda_1^2 t_r + \lambda_2^2 t_r + \lambda_3^2 t_r) + 3(\delta_1^2 t_r + \delta_2^2 t_r + \delta_3^2 t_r) + \\ & + 6(\delta_1 \delta_2 t_r + \delta_2 \delta_3 t_r) + \\ & + 6(\delta_1 \lambda_1 t_r + \delta_1 \lambda_2 t_r + \delta_2 \lambda_2 t_r + \delta_2 \lambda_3 t_r + \delta_3 \lambda_3 t_r) = \\ & = 3t_r (3\lambda^2 + 7\delta^2 + 10\lambda\delta). \end{aligned} \quad (2)$$

Формула (2) определяет интенсивность потока отказов исследуемой структуры.

Оценка имитационным моделированием.

Имитационная модель для структуры, представленной на рис. 1, может быть реализована сетью из следующих графов с независимыми состояниями:

V_1, V_2, V_3 — примитивные графы, моделирующие события отказов в соответствующих мажоритарных;

$A_1[1], A_1[2], A_1[3]$ — графы состояний экземпляров троированного блока A_1 ;

$A_2[1], A_2[2], A_2[3]$ — аналогично для блока A_2 ;

$A_3[1], A_3[2], A_3[3]$ — аналогично для блока A_3 ;

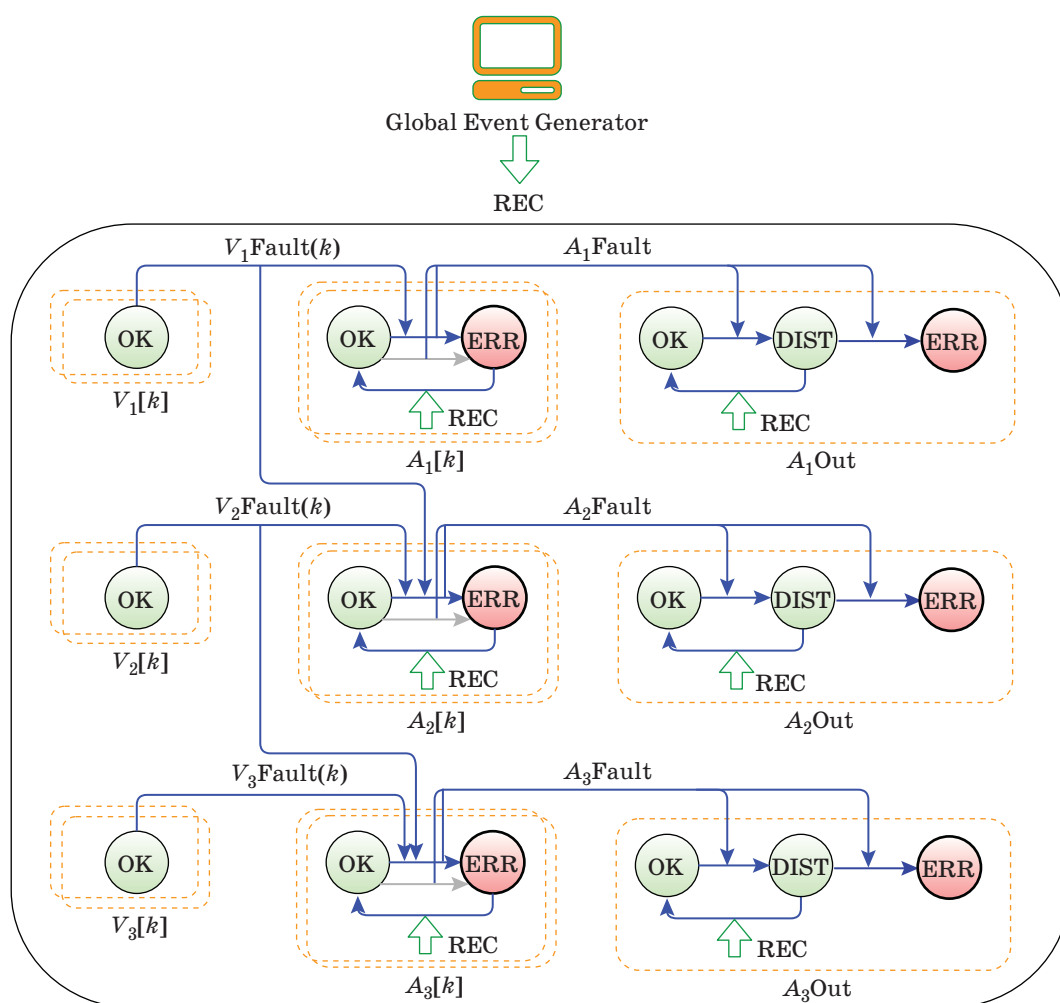
A_1Out, A_2Out, A_3Out — графы, моделирующие состояние выходных линий блоков O_1, O_2, O_3 .

Графическая интерпретация имитационной модели представлена на рис. 4.

Пунктирными линиями на рис. 4 обозначены независимые графы, входящие в состав сети (двойными пунктирными — тройки однотипных графов). Состояния, выделенные тонким контуром, являются работоспособными, жирным — неработоспособными. Если над переходом между состояниями нет названия, он обозначает внутреннее событие в графе. Если переход именованный, он обозначает внешнее событие, передаваемое между графами (стрелка, входящая во внутренний переход, символизирует порождение внутреннего перехода внешним событием). Стрелки, исходящие из перехода, иллюстрируют генерацию внешнего события при возникновении данного перехода. Если стрелка выходит из состояния, значит внутреннего перехода в исходном графе при возникновении события нет. Утолщенные стрелки обозначают воздействие глобального события, порождаемого генератором Global Event Generator независимо от переходов в каком-либо графе (в данном случае — событие восстановления REC (recovery) с постоянным периодом).

Система находится в работоспособном состоянии, если все троированные модули A_1, A_2, A_3 исправны, т. е. на их выходах выставлены корректные сигналы. Другими словами, логическая функция работоспособности системы $HealthFunction = A_1Out \& A_2Out \& A_3Out$.

Рассмотрим принцип функционирования модели на примере блока A_1 . Примитивные графы $V_1[1], V_1[2], V_1[3]$ моделируют работу подключенных к модулю мажоритаров, которые являются источниками отказов $V_1Fault[k]$, где $k = 1..3$. Как видно из рис. 4, данные отказы влияют на соответствующий граф $A_1[k]$, что переводит его в неработоспособное состояние ERR. Этот переход в свою очередь порождает событие A_1Fault . При этом в графе A_1Out , моделирующем состояние выходной линии тройки модулей A_1 , происходит переход в состояние DIST (distortion), свидетельствующее о том, что одна из выходных линий блока A_1 искажена. Отметим, что после этого граф $A_1[k]$, в котором произошел отказ, продолжает хранить неработоспособное состояние ERR и не может быть источником отказов для A_1Out . Возникновение события восстановления REC, воздействующего на все блоки, переводит в состояние ОК как $A_1[k]$, так и A_1Out . Отказ A_1Out произойдет в том случае, если событие A_1Fault возникнет дважды за период восстановления, другими словами, если два различных экземпляра блока A_1 перейдут в неработоспособное состояние



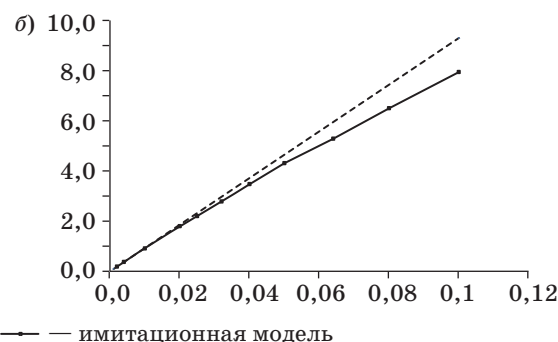
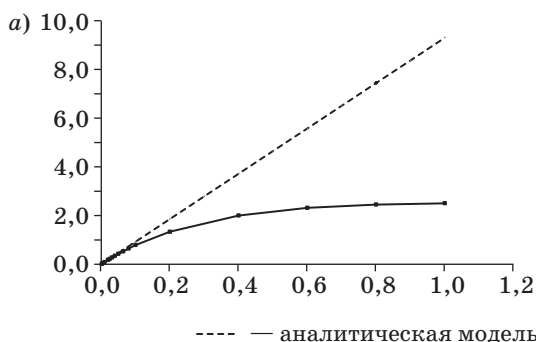
■ **Рис. 4.** Графическая интерпретация имитационной модели исследуемой системы

■ **Описание графов состояний сети, моделирующей анализируемую структуру**

Граф	Состояния	Переходы
$V_n[k]$ $n = 1..3$ — номер тройки мажоритаров $k = 1..3$ — номер экземпляра в тройке мажоритаров	ОК (работоспособное) — графы имеют всего одно состояние, так как непосредственно не влияют на работоспособность системы, а лишь являются источниками внешних сбоев для связанных модулей	ОК → ОК Закон распределения: экспоненциальный Интенсивность: $\delta = 0,01$ Возникает при отказе мажоритаров Генерируется событие $V_n \text{Fault}(k)$, свидетельствующее о сбое в соответствующем экземпляре мажоритаров k
$A_n[k]$ $n = 1..3$ — номер блока $k = 1..3$ — номер экземпляра в блоке	ОК (работоспособное) — соответствующий экземпляр блока исправен ERR (неработоспособное) — соответствующий экземпляр модуля неисправен вследствие возникновения восстанавливаемого отказа	ОК → ERR Закон распределения: экспоненциальный Интенсивность: $\lambda = 1,0$ Возникает при внутреннем отказе экземпляра блока Генерируется событие $A_n \text{Fault}$, свидетельствующее об отказе соответствующего экземпляра блока k ОК → ERR Возникает при генерации $V_n \text{Fault}(k)$ в одном из подключенных к модулю мажоритаров Генерируется событие $A_n \text{Fault}$ ERR → ОК Возникает при генерации глобального события REC

■ Окончание табл.

Граф	Состояния	Переходы
A_n Out $n = 1..3$ — номер блока	ОК (работоспособное) — троированный блок исправен	ОК → DIST Возникает при наступлении события A_n Fault
	DIST (работоспособное) — один из экземпляров блока исправен	DIST → ОК Возникает при генерации глобального события REC
	ERR (неработоспособное) — два или более экземпляра блока n неисправны, что приводит к возникновению невосстанавливаемого отказа и выходу системы из строя	DIST → ERR Возникает при наступлении события A_n Fault



■ Рис. 5. Зависимость интенсивности потока отказов системы от периода восстановления блоков A_1, A_2, A_3 : а — общий результат; б — начальный участок

в течение периода восстановления. Модель событий для блоков A_2 и A_3 строится аналогичным образом.

Более подробно связи между событиями и состояниями в графах описаны в таблице.

При расчете для каждого фиксированного значения t_r в диапазоне от 0,0 до 1,0 производится заданное число экспериментов (50 000), каждый из которых продолжается до перехода системы в неработоспособное состояние. Во время каждого эксперимента моделируются возникновение случайных и детерминированных событий и реакции системы, заключающиеся в переходах графов между состояниями, и производится измерение времени наработки до отказа. Итоговые результаты получают путем статистической оценки произведенного набора экспериментов.

Полученные результаты.

Расчет исследуемой структуры (рис. 5, а) проведен с помощью комбинаторной оценки и рассчитанной имитационной модели. Начальный участок данных кривых показан на рис. 5, б.

Ось абсцисс: период восстановления блоков A_1, A_2, A_3 ($T_{\text{восст}} = 1/\lambda_r$).

Ось ординат: интенсивность потока отказов на выходах O_1, O_2, O_3 . Относительная единица измерения 1,0 обозначает интенсивность сбоя в любом из троированных элементов A_1, A_2, A_3 . Интенсивность сбоев в любом из троированных мажоритаров равна 0,01.

Закключение

Полученные результаты позволяют сделать следующие выводы.

1. Аналитическая формула дает достаточно точную оценку надежности анализируемой структуры для малых периодов восстановления. Из рис. 5, б видно, что при периоде восстановления меньше 0,1 (в десять раз меньшем приближительного периода между сбоями в каждом блоке) обе характеристики имеют линейную зависимость с одинаковым коэффициентом наклона.

2. Если обеспечение достаточно малого периода восстановления блоков невозможно из-за недостаточного резерва времени, зависимость интенсивности отказов от периода восстановления выходит на нелинейный участок, что приводит к необходимости применять имитационную модель для корректного расчета.

3. Расчет с помощью имитационной модели позволяет оценить эффективность применения механизмов повышения отказоустойчивости:

— если бы в системе не было реализовано ни восстановление, ни структурное резервирование, то отказ в любом из модулей A_1, A_2, A_3 приводил бы к выходу из строя всей системы. С учетом того, что периоды между отказами являются случайной величиной, подчиненной экспоненциальному закону распределения, а интенсивность возникновения отказов в каждом блоке равна 1,0,

среднее время наработки до отказа может быть оценено как $T_{исх} = \frac{1}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{1}{3 \cdot 1,0} = 0,333$;

— если применяется троирование блоков и мажоритаров (что соответствует результатам расчета имитационной модели на больших периодах восстановления), то при интенсивности отказов каждого мажоритары, равной 0,01, интенсивность отказов системы стремится приблизительно к 2,5, а время средней наработки до отказа — к $T_{рез} = \frac{1}{2,5} = 0,4$ (эти величины соответствуют

результату расчета марковской модели при отсутствии переходов по событию восстановления). Таким образом, по сравнению с нерезервированной системой время наработки до отказа было увеличено на $\frac{T_{рез} - T_{исх}}{T_{исх}} = \frac{0,4 - 0,333}{0,333} \cdot 100\% \approx 20\%$;

— обеспечение регулярного восстановления искаженных данных с периодом 0,05 (в 20 раз чаще возникновения сбоев в модулях) приводит к увеличению периода наработки до отказа до $T_{восст} = 1/0,43 = 2,33$, т. е. приблизительно в 5 раз относительно резервированной, но не восстанавливаемой структуры, что доказывает эффектив-

ность применения восстановления для борьбы с «мягкими» отказами.

4. Большое влияние периода восстановления на надежность системы подтверждает необходимость учета программного обеспечения при проектировании системы, так как оно определяет допустимый резерв времени, отводимый на периодическое восстановление блоков. Так как имитационная модель пригодна для анализа при любых величинах периода восстановления, она помогает оценить, достаточен ли имеющийся резерв времени для обеспечения требуемых характеристик надежности.

5. Имитационная модель также необходима для сравнительного анализа возможных организаций восстанавливаемых блоков на этапе синтеза с учетом таких показателей, как сложность реализации и быстродействие.

6. После того как в процессе проектирования были определены параметры восстановления для каждого блока и с помощью имитационной модели вычислена интенсивность потока отказов на выходе блоков, надежность сети блоков может быть оценена с помощью традиционных аналитических моделей или логико-вероятностными методами.

Литература

1. Edmonds L. D., Barnes C. E., Scheick L. Z. An Introduction to Space Radiation Effects on Microelectronics // JPL Publication 00-06, 2000. <http://snebulos.mit.edu/projects/reference/NASA-Generic/JPL-00-06.pdf> (дата обращения: 05.12.2015).
2. Gaillard R. Single Event Effects Mechanisms and Classification // Frontiers in Electronic Testing. 2011. Vol. 41. P. 27–54.
3. Amusan O. A., et al. Single Event Upsets in Deep-Submicrometer Technologies due to Charge Sharing/ O. A. Amusan, L. W. Massengill, M. P. Baze, A. L. Sternberg, A. F. Witulski, B. L. Bhuvu, J. D. Black // IEEE Transactions on Device and Materials Reliability. 2008. Vol. 8. N 3. P. 582–589.
4. James R. Schwank, Marty R. Shaneyfelt, Paul E. Dodd. Radiation Hardness Assurance Testing of Microelectronic Devices and Integrated Circuits: Radiation Environments, Physical mechanisms, and Foundations for Hardness Assurance // IEEE Transactions on Nuclear Science. 2013. Vol. 60. N 3. P. 2074–2100.
5. Максименко С. Л., Мелехин В. Ф., Филиппов А. С. Анализ проблемы построения радиационно-стойких информационно-управляющих систем // Информационно-управляющие системы. 2012. № 2(57). С. 18–25.
6. Максименко С. Л., Мамутова О. В., Филиппов А. С., Мелехин В. Ф. Методология проектирования вос-

станавливаемых встраиваемых вычислительных систем // Университетский научный журнал. 2014. № 8. С. 144–153.

7. Jacob A. Abraham, Daniel P. Siewiorek. An Algorithm for the Accurate Reliability Evaluation of Triple Modular Redundancy Networks // IEEE Transactions on Computers. 1974. Vol. C-23. N 7. P. 682–692.
8. Максименко С. Л., Мелехин В. Ф. Анализ надежности функциональных узлов цифровых СБИС со структурным резервированием и периодическим восстановлением работоспособного состояния // Информационно-управляющие системы. 2013. № 2(63). С. 18–23.
9. Черкесов Г. Н. Надежность аппаратно-программных комплексов. — СПб.: Питер, 2005. — 480 с.
10. Черкесов Г. Н., Можжаев А. С. Логико-вероятностные методы расчета надежности структурно-сложных систем // Качество и надежность изделий. Вып. 3 (15). — М.: Знание, 1991. С. 3–65.
11. Черкесов Г. Н., Степанов Ю. В. Логико-вероятностный анализ надежности сложных систем на основе общего решения систем логических уравнений // Научно-технические ведомости СПбГТУ. 2003. № 2. С. 149–158.
12. Мельников И. В. Применение имитационной модели надежности при проектировании изделий ракетно-космической техники // Молодой ученый. 2011. № 9. С. 39–42.

UDC 681.3

doi:10.15217/issn1684-8853.2016.2.26

Methods and Tools for Structural Block Reliability Analysis with Reservation and Periodic Information Recovery at Various Stages of Computing System DesignEgorov I. V.^a, Post-Graduate Student, iegorov@kspt.icc.spbstu.ruMelekhin V. F.^a, Dr. Sc., Tech., Professor, melekhin@kspt.ftk.spbstu.ru^aPeter the Great St. Petersburg Polytechnic University, 29, Politekhnikheskaia St., 195251, Saint-Petersburg, Russian Federation

Introduction: When designing fault-tolerant computing systems which must work under radiation, you have to estimate the influence of information distortion caused by charged particles. Software tools developed to analyze the reliability of such systems should take into account these particle-caused failures and simulate the features of periodic recovery of the distorted information which are used to increase the reliability. **Purpose:** The goal of this study is to determine the bounds within which you can apply the reliability analysis methods based on using analytical and imitational models in order to evaluate the reliability of systems with periodic recovery. Another goal is to compare the results of analyzing the reliability characteristics obtained by the analytical and imitational models. **Results:** The analytical and imitating models helped to calculate and analyze the time for which a recoverable system works until a failure happens. It has been found out that determined recovery processes restrict the applicability of the traditional reliability estimation approaches such as Markov models or logical-probabilistic methods. The necessity to take into account the software influence when designing systems with periodic recovery has been substantiated. **Practical relevance:** The conducted research helped to specify the application areas for certain reliability models at various stages of designing fault-tolerant recoverable computing systems.

Keywords — Imitational Modeling, Reliability of Recoverable Systems, Radiation Resistant Computing System, Markov Model, Structural Reservation, Information Redundancy.

References

- Edmonds L. D., Barnes C. E., Scheick L. Z. An Introduction to Space Radiation Effects on Microelectronics. JPL publication, 2000, no. 00-06. Available at: <http://snebulos.mit.edu/projects/reference/NASA-Generic/JPL-00-06.pdf> (accessed 05 December 2015).
- Gaillard R. Single Event Effects Mechanisms and Classification. *Frontiers in Electronic Testing*, 2011, vol. 41, pp. 27–54.
- Amusan O. A., Massengill L. W., Baze M. P., Sternberg A. L., Witulski A. F., Bhuvu B. L., Black J. D. Single Event Upsets in Deep-Submicrometer Technologies due to Charge Sharing. *IEEE Transactions on Device and Materials Reliability*, 2008, vol. 8, no. 3, pp. 582–589.
- James R. Schwank, Marty R. Shaneyfelt, Paul E. Dodd. Radiation Hardness Assurance Testing of Microelectronic Devices and Integrated Circuits: Radiation Environments, Physical mechanisms, and Foundations for Hardness Assurance. *IEEE Transactions on Nuclear Science*, 2013, vol. 60, no. 3, pp. 2074–2100.
- Maximenko S. L., Melekhin V. F., Filippov A. S. Analysis of the Problem of Radiation-Tolerant Information and Control-Systems Implementation. *Informatsionno-upravliaushchie sistemy* [Information and Control Systems], 2012, no. 2(57), pp. 18–25 (In Russian).
- Maximenko S. L., Filippov A. S., Melekhin V. F., Mamoutova O. V. Design Methodology for Embedded Systems with Built-in Self-Recovery. *Universitetskii nauchnyi zhurnal*, 2014, no. 8, pp. 144–153 (In Russian).
- Jacob A. Abraham, Daniel P. Siewiorek. An Algorithm for the Accurate Reliability Evaluation of Triple Modular Redundancy Networks. *IEEE Transactions on Computers*, 1974, vol. C-23, no. 7, pp. 682–692.
- Maximenko S. L., Melekhin V. F. Analysis of Reliability of Functional Nodes of Digital VLSI Circuits with Structural Redundancy and Periodic Operational State Recovery. *Informatsionno-upravliaushchie sistemy* [Information and Control Systems], 2013, no. 2(63), pp. 18–23 (In Russian).
- Cherkesov G. N. *Nadezhnost' apparatno-programmnykh kompleksov* [Reliability of Hardware and Software Systems]. Saint-Petersburg, Piter Publ., 2005. 480 p. (In Russian).
- Cherkesov G. N., Mozhaev A. S. Logical-and-Probabilistic Methods of Calculation of Reliability of Structural and Difficult Systems. *Kachestvo i nadezhnost' izdelii*, 1991, no. 3(15), pp. 3–65 (In Russian).
- Cherkesov G. N., Stepanov Y. V. Logical-and-Probabilistic Analysis of Reliability of Difficult Systems on the Basis of the Common Decision of Systems of the Logical Equations. *Nauchno-tehnicheskie vedomosti SPbGTU*, 2003, no. 2, pp. 149–158 (In Russian).
- Melnikov I. V. Application of Imitating Model of Reliability at Design of Products of the Missile and Space Equipment. *Molodoj uchenyj*, 2011, no. 9, pp. 39–42 (In Russian).

УДК 681.5

doi:10.15217/issn1684-8853.2016.2.35

НЕЧЕТКОЕ УПРАВЛЕНИЕ АВТОМОБИЛЬНОЙ АНТИБЛОКИРОВОЧНОЙ СИСТЕМОЙ

М. В. Бураков^а, канд. техн. наук, доцент**А. С. Коновалов^а**, доктор техн. наук, профессор^аСанкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ

Постановка проблемы: тормозные антиблокировочные системы используются в современных автомобилях для предотвращения юза колес при торможении. Модель транспортного средства имеет нелинейную форму. Регулятор должен обеспечивать контролируемый вращающий момент, необходимый для поддержания оптимального значения коэффициента проскальзывания колеса. **Цель:** построение генетически настраиваемого нечеткого регулятора ПИД-типа для повышения качества работы антиблокировочной системы по сравнению с линейными ПИД-регуляторами. **Результаты:** разработана антиблокировочная система с использованием модели колеса автомобиля и тормозного привода при движении транспортного средства в продольном направлении. Для управления торможением предложены три типа контроллера: релейный регулятор, генетически настраиваемые линейный ПИД и нечеткий регулятор ПИД-типа. Качество работы антиблокировочной системы оценивается в системе MatLab по тормозному пути и коэффициенту продольного скольжения автомобиля. Нечеткий регулятор показал лучшие характеристики для модели антиблокировочной системы, уменьшая тормозной путь до 10 % по сравнению с обычным ПИД и более чем на 30 % по сравнению с релейным контроллером. **Практическая значимость:** предложенный алгоритм управления перспективен для реализации в антиблокировочной системе в режиме реального времени.

Ключевые слова — антиблокировочная система, нечеткий логический регулятор, ПИД-регулятор, генетический алгоритм.

Введение

В последние годы при разработке систем управления сложными техническими объектами все более широко используется нечеткое логическое управление. Одной из трудных практических задач, где использование нечетких логических регуляторов (НЛР) показало хорошую эффективность, является задача конструирования автомобильных антиблокировочных систем (АБС).

Разработка эффективных АБС является одним из основных условий повышения безопасности эксплуатации автомобильного транспорта. Экстренное задействование тормозов с целью остановить автомобиль либо снизить его скорость может привести к противоположному результату — колеса блокируются и теряют сцепление с дорожным покрытием, а автомобиль сохраняет скорость и перестает слушаться руля. Первые АБС, которые появились в автомобильной промышленности в 60-е гг. XX в., построенные на аналоговых процессорах, были дорогостоящими и ненадежными [1, 2]. Современные встроенные автомобильные микропроцессорные системы позволяют реализовывать сложные алгоритмы управления, в том числе на базе нечеткой логики.

Основу НЛР составляют нечеткие правила; они связывают наблюдаемое состояние объекта и управление, которое должно в этом состоянии

использоваться. Таким образом, можно рассматривать НЛР как разновидность экспертных систем, в которых знания имеют явное представление в виде правил. При проектировании АБС также используется такой подход. Набор нечетких правил позволяет описать нелинейный закон управления и повысить качество работы АБС ([3–5] и др.). Однако более широкие возможности обеспечивает поисковая оптимизация НЛР, поскольку человек, в силу своих ограниченных психофизических возможностей, является ненадежным источником информации при анализе быстропротекающих процессов в момент торможения автомобиля.

Сложность задачи оптимизации НЛР требует применения методов глобального поиска, таких как генетический алгоритм или метод роя частиц [6–8]. Эти методы используют целевую функцию как невязку между заданным эталонным процессом и выходом имитационной модели при конкретном наборе параметров. В процессе оптимизации меняются параметры регулятора, а параметры модели остаются неизменными. Таким образом, чем меньше настраиваемых параметров имеет НЛР, тем проще задача оптимизации. В данной работе при конструировании АБС рассматривается генетическая оптимизация нечеткого регулятора ПИД-типа, который представляет собой 3-канальную структуру с кусочно-линейной аппроксимацией нелинейной функции каждого канала [9–11].

Математическое описание процесса торможения

При построении описания АБС автомобиля обычно используют следующие допущения [2]:

- динамика колес автомобиля идентична;
- масса автомобиля равномерно распределяется по всем четырем колесам;
- влияние трансмиссии и подвески автомобиля не учитывается.

Таким образом, может быть рассмотрена одноколесная модель автомобиля в процессе торможения (рис. 1). На рисунке приняты обозначения: M_T — тормозной момент, Н·м; F_x — продольная составляющая контактной силы колеса, Н; v — абсолютная скорость автомобиля, м/с; ω — угловая скорость колеса, рад/с; F_N — сила реакции опоры (нормальная сила), Н.

Уравнения движения автомобиля имеют вид

$$\begin{cases} J \frac{d\omega(t)}{dt} = F_x(t)r - M_T(t); \\ m\dot{v}(t) = -F_x(t) \end{cases} \quad (1)$$

где J — момент инерции колеса, кг·м²; r — радиус колеса, м.

Продольная составляющая контактной силы определяется выражением

$$F_x = \mu(s(t))F_N, \quad (2)$$

где μ — коэффициент дорожного трения; s — продольное проскальзывание.

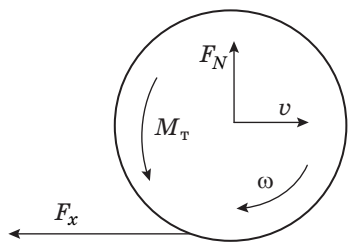
Нормальная сила получается по формуле

$$F_N = mg, \quad (3)$$

где m — приведенная на одно колесо масса автомобиля ($m = M/4$, где M — масса автомобиля); g — ускорение свободного падения.

Проскальзывание колеса определяется формулой

$$s(t) = \frac{v_s(t)}{v(t)} = \frac{v(t) - v_\omega(t)}{v(t)} = \frac{v(t) - \omega(t)r}{v(t)}. \quad (4)$$



■ Рис. 1. Модель тормозящего колеса

Таким образом:

$$\omega(t) = \frac{v(t)}{r}(1 - s(t)).$$

При $s=0$ наблюдается идеальное сцепление с дорогой, полное отсутствие скольжения. При $s=1$ колесо заблокировано, происходит неуправляемое скольжение.

На величину $\mu(s)$ влияют многие факторы: состояние дороги, погодные условия, состояние шины, скорость автомобиля. Примерные зависимости $\mu(s)$ для различных состояний дорожного покрытия приведены на рис. 2.

Может быть использовано аналитическое описание коэффициента дорожного трения с помощью формулы [12]

$$\mu(s) = a(1 - e^{-bs} - cs), \quad (5)$$

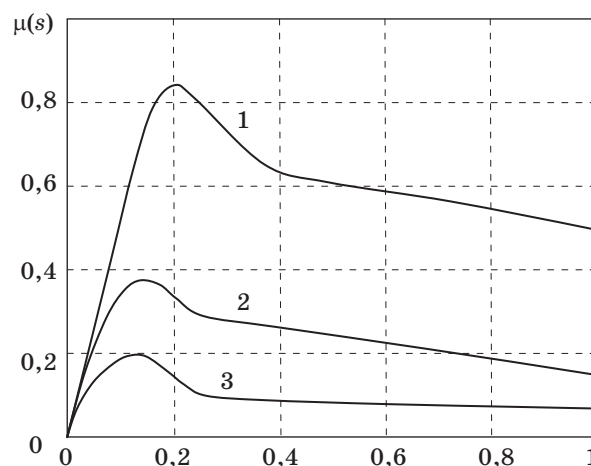
где a , b и c — коэффициенты, зависящие от состояния дорожного покрытия.

Например, для сухого асфальта значения коэффициентов: $a = 1,28$, $b = 23,99$ и $c = 0,52$.

Для вычисления значения проскальзывания нужно иметь информацию об угловой и линейной скорости колеса. Измерение угловой скорости колеса с помощью датчиков является достаточно простой задачей. Однако точно измерить абсолютную скорость автомобиля в реальном времени сложно, что приводит к необходимости использовать оценки проскальзывания.

В работе [13] для описания коэффициента дорожного трения использована формула, учитывающая скорость движения автомобиля:

$$\mu(s, v) = (C_1(1 - e^{-C_2s}) - C_3s)e^{-C_4sv}. \quad (6)$$



■ Рис. 2. Зависимость коэффициента дорожного трения от продольного проскальзывания; дорожное покрытие: 1 — сухое, 2 — мокрое, 3 — скользкое (обледенелое)

■ Параметры для расчета коэффициента дорожного трения

Дорожное полотно	C_1	C_2	C_3	C_4
Сухой асфальт	1,029	17,16	0,523	0,03
Сухой бетон	1,1973	25,168	0,5373	0,03
Снег	0,1946	94,129	0,0646	0,03
Лед	0,05	306,39	0	0,03

Параметры, входящие в (6), приведены в таблице.

Рассмотрим производную продольного проскальзывания:

$$\begin{aligned}\dot{s}(t) &= -\frac{r}{v(t)}\dot{\omega}(t) + \frac{r\omega(t)}{v^2(t)}\dot{v}(t) = \\ &= \frac{-r\dot{\omega}(t) + (1-s(t))\dot{v}(t)}{v(t)}.\end{aligned}\quad (7)$$

Из (1)–(3) следует

$$\dot{\omega}(t) = \frac{mg\mu(s(t))r - M_T(t)}{J}.\quad (8)$$

Подставляя (8) в (7), можно записать систему уравнений, описывающих динамику колеса при торможении:

$$\begin{cases}\dot{s}(t) = -\frac{1}{v(t)}\left(\frac{1-s(t)}{m} + \frac{r^2}{J}\right)mg\mu(s(t)) + \\ + \frac{r}{v(t)J}M_T(t); \\ \dot{v}(t) = -g\mu(s(t)).\end{cases}\quad (9)$$

Тормозной момент может быть описан упрощенно:

$$M_T(t) = \begin{cases} kP(t), & kP(t) < M_{\max}; \\ M_{\max}, & kP(t) \geq M_{\max}, \end{cases}\quad (10)$$

где k — константа торможения; P — давление, создаваемое тормозной системой при прижатии колодки к тормозному диску; $M_{\max}$ — максимальное давление в тормозной системе.

Более реалистично тормозной момент можно описать передаточной функцией с запаздыванием [14]:

$$\frac{M_T(s)}{P(s)} = e^{-\tau s} \frac{ka}{s+a}.\quad (11)$$

Уравнения (5), (6) и (9)–(11) представляют собой модель динамики автомобиля при торможении. При проведении вычислительных экспериментов будем рассматривать легковой автомо-

биль со следующими параметрами: начальная скорость $v_0 = 100$ км/ч; масса (приведенная на одно колесо) $m = 375$ кг; радиус колеса $r = 0,32$ м; момент инерции колеса $J = 1,7$ кг·м²; максимальный тормозной момент 2500 Н·м.

Модель антиблокировочной системы

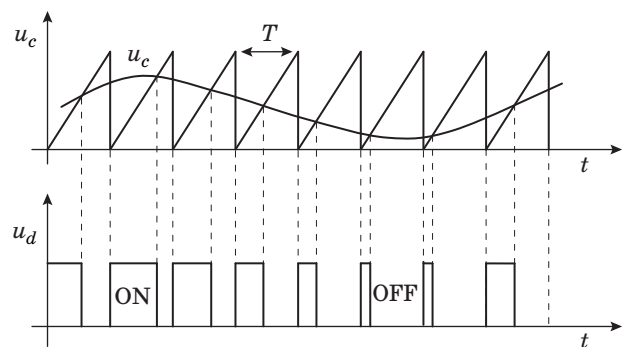
Типичная АБС состоит из центрального микропроцессора, четырех датчиков скорости вращения колес (по одному на каждое колесо) и двух или четырех гидравлических или пневматических клапанов в цепи управления тормозами.

В основу работы колесных датчиков положен принцип электромагнитной индукции. При вращении колеса мимо датчика проходят зубцы и впадины специального ротора и наводят в обмотке датчика электрический сигнал, частота которого пропорциональна угловой скорости колеса и количеству зубцов на роторе.

Для сохранения прямолинейного движения автомобиля после начала торможения микропроцессор анализирует скорость вращения каждого колеса. Если какое-то колесо вращается намного медленнее других, то происходит уменьшение давления в соответствующем тормозном цилиндре.

Для управления клапанами может быть использована широтно-импульсная модуляция (ШИМ), вырабатывающая с периодом T пилообразный сигнал (рис. 3), с которым сравнивается текущее значение сигнала управления u_c . В результате сравнения клапан тормозной системы переключается в режим торможения (ON) или в режим растормаживания (OFF).

Гидравлическая тормозная система, а также микропроцессор создают некоторую временную задержку, поэтому для вычисления ошибки неверно использовать значение коэффициента скольжения, получаемое на выходе системы. Необходимо в реальном времени предсказывать будущее значение коэффициента скольжения и именно его использовать для контроля ошибки [15].



■ Рис. 3. Использование ШИМ

Модель наблюдателя коэффициента скольжения можно построить, записав (4) в виде

$$s(t)v(t) = v(t) - r\omega(t). \quad (12)$$

Дифференцируя (12) по времени, получаем

$$\dot{s}(t) = \frac{1}{v(t)}((1-s(t))\dot{v}(t) - r\dot{\omega}(t)). \quad (13)$$

Заменим производные конечными разностями:

$$\begin{aligned} \dot{s}(t) &\approx \frac{s(k+1) - s(k)}{\Delta t}; & \dot{v}(t) &\approx \frac{v(k+1) - v(k)}{\Delta t}; \\ \dot{\omega}(t) &\approx \frac{\omega(k+1) - \omega(k)}{\Delta t}, \end{aligned}$$

где k — момент времени; Δt — интервал дискретизации.

Тогда выражение (13) преобразуется в формулу для предсказания коэффициента скольжения:

$$\begin{aligned} s(k+1) &= s(k) + \frac{\Delta t}{v(k)} \times \\ &\times \left((1-s(k)) \left(\frac{v(k+1) - v(k)}{\Delta t} \right) - r \left(\frac{\omega(k+1) - \omega(k)}{\Delta t} \right) \right). \end{aligned}$$

Модель АБС приобретает вид, показанный на рис. 4.

При моделировании сравнивались три варианта закона управления.

1. Релейное управление:

$$u_c(t) = \begin{cases} P_{\max}, & e(t) > 0; \\ 0, & e(t) \leq 0. \end{cases}$$

2. ПИД-регулятор с насыщением:

$$u_c(t) = \begin{cases} u = k_p e(t) + k_d \frac{de(t)}{dt} + k_i \int e(t) dt, & u < P_{\max}; \\ P_{\max}, & u \geq P_{\max}; \\ 0, & e(t) \leq 0. \end{cases}$$

3. НЛР с насыщением:

$$u_c(t) = \begin{cases} u = F_p(e(t)) + F_d\left(\frac{de(t)}{dt}\right) + F_i\left(\int e(t) dt\right), & u < P_{\max}; \\ P_{\max}, & u \geq P_{\max}; \\ 0, & e(t) \leq 0. \end{cases}$$

Здесь k_p, k_d, k_i — настраиваемые коэффициенты ПИД-регулятора; F_p, F_d, F_i — настраиваемые нелинейные функции, аппроксимирующие нечеткий закон управления [9–11].

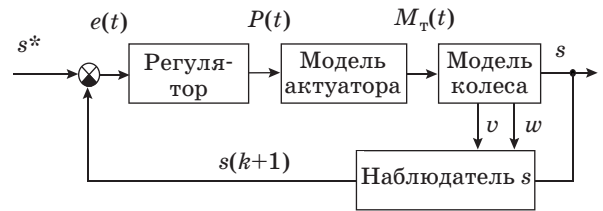


Рис. 4. Модель АБС с наблюдателем коэффициента скольжения

Для настройки ПИД-регулятора и НЛР ПИД-типа использовался генетический алгоритм. Алгоритм генетической настройки НЛР описан в работе [11].

Результаты моделирования

Блок-схема эксперимента в MatLabSimulink представлена на рис. 5. Исследовались три варианта управления: релейный регулятор, ПИД-регулятор и НЛР ПИД-типа.

Переходные процессы изменения скорости автомобиля при торможении для различных регуляторов показаны на рис. 6, а–в.

Структура НЛР ПИД-типа после обучения представлена на рис. 7. Фильтр низких частот на выходе регулятора служит для сглаживания резких скачков сигнала управления.

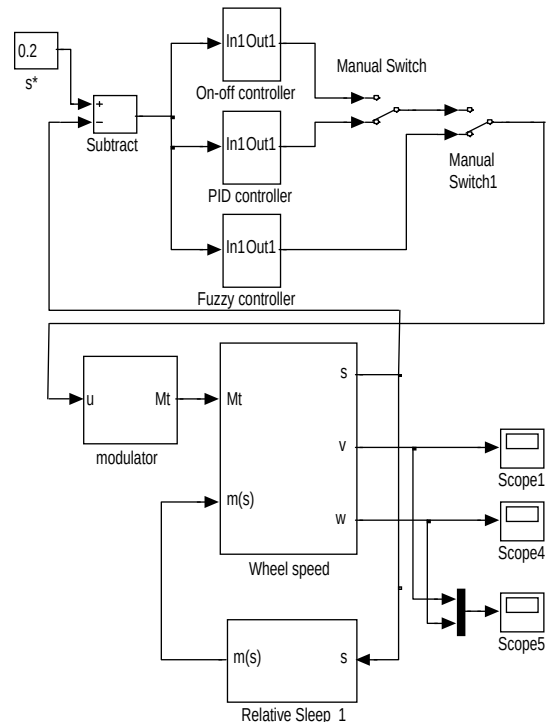
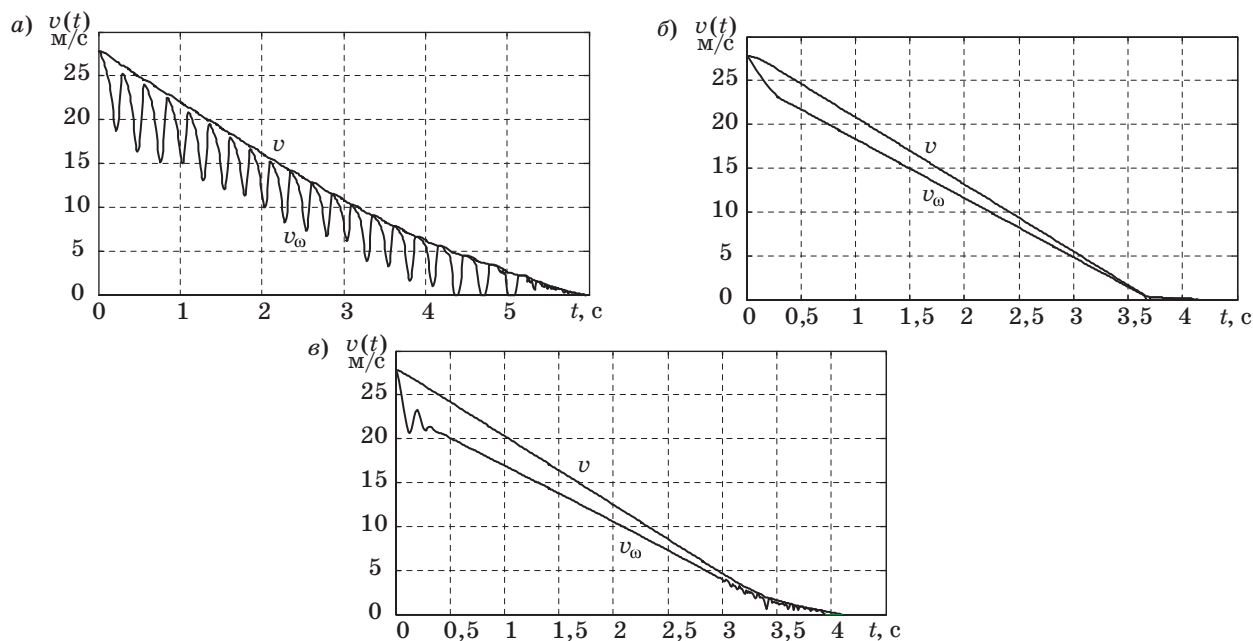


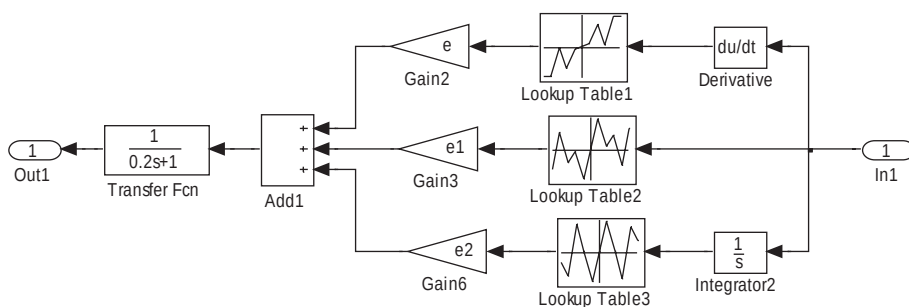
Рис. 5. Исследование различных законов управления в АБС

Сравнение сигналов управления, вырабатываемых ПИД-регулятором и НЛР во время переходных процессов, показано на рис. 8, а на рис. 9 — тормозной путь при различных законах управления.

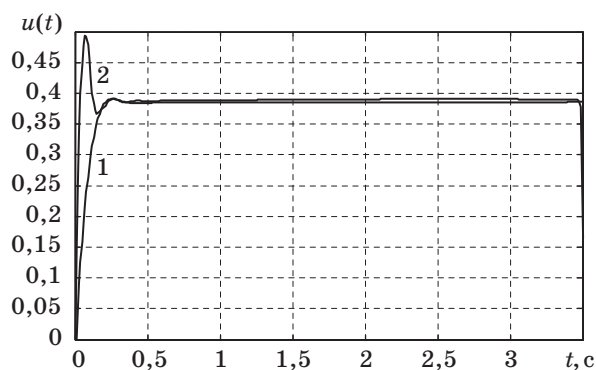
Таким образом, результаты моделирования демонстрируют значительное преимущество нелинейного НЛР ПИД-типа по сравнению с альтернативными вариантами регуляторов.



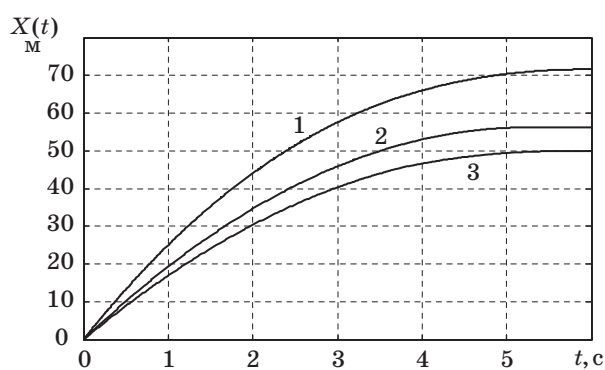
■ Рис. 6. Переходный процесс при релейном законе управления (а); под управлением ПИД-регулятора (б); под управлением НЛР (в)



■ Рис. 7. Нечеткий регулятор ПИД-типа после обучения



■ Рис. 8. Сигналы управления: 1 — ПИД-регулятор; 2 — НЛР ПИД-типа



■ Рис. 9. Тормозной путь: 1 — релейный регулятор; 2 — ПИД-регулятор; 3 — НЛР ПИД-типа

Заключение

При моделировании работы АБС рассматривалась одноколесная модель автомобиля, основными параметрами которой являются масса автомобиля, радиус и момент инерции колеса, а также максимальный тормозной момент. Условия движения автомобиля по конкретному дорожному покрытию задаются зависимостью коэффициента дорожного трения (коэффициента скольжения) от проскальзывания колеса $\mu(s)$. Эта зависимость может быть задана аналитически для разных типовых состояний дорожного покрытия. Для более точного описания коэффициента скольжения необходимо учитывать скорость движения автомобиля.

Поскольку гидравлическая тормозная система и микропроцессор создают некоторую временную задержку, для вычисления ошибки неверно

использовать текущее значение коэффициента скольжения. Необходимо в реальном времени предсказывать будущее значение коэффициента скольжения с помощью наблюдателя коэффициента скольжения.

Исследовались три варианта управления: релейный регулятор, а также генетически настраиваемый линейный ПИД-регулятор и нелинейный НЛР ПИД-типа. Результаты вычислительных экспериментов показали, что регулятор релейного типа значительно уступает НЛР и ПИД-регуляторам по длине тормозного пути. НЛР обеспечивает сокращение тормозного пути по сравнению с линейным ПИД-регулятором примерно на 10 % и более чем на 30 % по сравнению с релейным контроллером.

Предложенный алгоритм нечеткого управления перспективен для реализации в системе АБС в режиме реального времени.

Литература

1. Ayman A., El-Shafei Z., Ahmed H., Farhan S. An Antilock-Braking Systems (ABS) Control: A Technical Review // *Intelligent Control and Automation*. 2011. N 2. P. 186–195.
2. Savaresi S. M., Tanelli M. Active Braking Control Systems Design for Vehicles. — London: Springer-Verlag, 2010. — 255 p.
3. Mauer G. F. A Fuzzy Logic Controller for an ABS Braking System // *IEEE Transactions on Fuzzy Systems*. 1995. Vol. 3. N 4. P. 381–388.
4. Will A. B., Zak S. H. Antilock Brake System Modelling and Fuzzy Control // *Intern. Journal of Vehicle Design*. 2000. Vol. 24. N 1. P. 1–18.
5. Precup R. E., Preitl S., Balas M., Balas V. Fuzzy Controllers for Tire Slip Control in Anti-lock Braking Systems // *IEEE Intern. Conf. on Fuzzy Systems*. Budapest. 2004. P. 1317–1322.
6. Burakov M. V., Konovalov A. S. Peculiarities of Genetic Algorithm Usage when Synthesizing Neural and Fuzzy Regulators // *Kluwer Intern. Series in Engineering and Computer Science*. 2002. N 664. P. 39–48.
7. Бураков М. В. Генетический алгоритм: теория и практика. — СПб.: ГУАП, 2008. — 164 с.
8. Yang X. S. Engineering Optimization: An Introduction with Metaheuristic Applications. — Hoboken: John Wiley & Sons, 2010. — 347 p.
9. Бураков М. В., Кирпичников А. П. Нечеткий регулятор ПИД-типа для нелинейного объекта // *Вестник Казанского технологического университета*. 2015. Т. 18. № 4. С. 242–244.
10. Бураков М. В., Яковец О. Б. Нечеткое управление силовым гироскопическим прибором // *Изв. вузов. Приборостроение*. 2015. Т. 58. № 10. С. 157–166.
11. Бураков М. В., Коновалов А. С., Яковец О. Б. Эволюционный синтез нечетких регуляторов // *Информационно-управляющие системы*. 2015. № 6. С. 28–33. doi:10.15217/issn1684-8853.2015.6.28
12. Pacejka H. B., Besselink I. Y. Magic Formula Tyre Model with Transient Properties // *Supplement to Vehicle System Dynamics*. 1997. Vol. 27. P. 234–249.
13. Oudghiri M., Chadli M., El Hajjaji A. Robust Fuzzy Sliding Mode Control for Antilock Braking System // *Intern. Journal on Sciences and Techniques of Automatic Control*. 2007. N 1. P. 13–28.
14. Chankit J., Rahul A., Abhishek D. Linear Control Technique for Anti-Lock Braking System // *Intern. Journal of Engineering Research and Applications*. 2014. Vol. 4. N 8. P. 104–108.
15. Chen C. K., Shin M. C. PID-Type Fuzzy Control for Anti-Lock Brake Systems with Parameter Adaptation // *JSME Intern. Journal*. 2004. Vol. 47. N 2. P. 675–685.

UDC 681.5

doi:10.15217/issn1684-8853.2016.2.35

Fuzzy Control for Anti-Lock Automobile Brake SystemsBurakov M. V.^a, PhD, Tech., Associate Professor, bmv@sknt.ruKonovalov A. S.^a, Dr. Sc., Tech., Professor, ask@aanet.ru^aSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

Introduction: Anti-lock braking systems are used in modern cars to prevent the wheels from locking after brakes were applied. A vehicle model has a nonlinear form. The controller should provide a controlled torque necessary to maintain the optimum value of the wheel slip ratio. **Purpose:** The goal of this study is building genetically adjusted fuzzy PID controllers to improve the control performance of anti-lock braking systems compared to the conventional PID controllers. **Results:** An anti-lock braking system has been developed which uses a quarter vehicle model and a brake actuator. The vehicle model is derived and simulated in the longitudinal direction. Three types of controller are proposed for building of anti-lock braking systems: a bang-bang controller, a linear PID and a fuzzy PID-type controller (the two latter ones are genetically adjusted). The system performance is evaluated with MatLab by the stopping distance and longitudinal slip of the vehicle. The fuzzy logic controller has shown the best performance for the anti-lock braking system model, reducing the stopping distance up to 10% compared to the conventional PID and over 30% compared to the bang-bang controller. **Practical relevance:** The control algorithm proposed in this paper has great potentials for its implementation in real-time anti-lock braking systems.

Keywords — Anti-lock Braking System, Fuzzy Logic Controller, PID Controller, Genetic Algorithm.

References

1. Ayman A., El-Shafei Z., Ahmed H., Farhan S. An Antilock-Braking Systems (ABS) Control: A Technical Review. *Intelligent Control and Automation*, 2011, no. 2, pp. 186–195.
2. Savaresi S. M., Tanelli M. *Active Braking Control Systems Design for Vehicles*. London, Springer-Verlag, 2010. 255 p.
3. Mauer G. F. A Fuzzy Logic Controller for an ABS Braking System. *IEEE Transactions on Fuzzy Systems*, 1995, vol. 3, no. 4, pp. 381–388.
4. Will A. B., Zak S. H. Antilock Brake System Modelling and Fuzzy Control. *Intern. Journal of Vehicle Design*, 2000, vol. 24, no. 1, pp. 1–18.
5. Precup R. E., Preitl S., Balas M., Balas V. Fuzzy Controllers for Tire Slip Control in Anti-lock Braking Systems. *IEEE Intern. Conf. on Fuzzy Systems*, Budapest, 2004, pp. 1317–1322.
6. Burakov M. V., Konovalov A. S. Peculiarities of Genetic Algorithm Usage when Synthesizing Neural and Fuzzy Regulators. *Kluwer Intern. Series in Engineering and Computer Science*, 2002, no. 664, pp. 39–48.
7. Burakov M. V. *Geneticheskii algoritm: Teoriia i praktika* [Genetic Algorithms: Theory and Practice]. Saint-Petersburg, GUAP Publ., 2008. 164 p. (In Russian).
8. Yang X. S. *Engineering Optimization: An Introduction with Metaheuristic Applications*. Hoboken, NJ, John Wiley & Sons, 2010. 347 p.
9. Burakov M. V., Kirpichnikov A. P. Fuzzy PID Controller for Nonlinear Plant. *Vestnik Kazanskogo tekhnologicheskogo universiteta*, 2015, vol. 18, no. 4, pp. 242–244 (In Russian).
10. Burakov M. V., Yakovets O. B. Fuzzy Control of Power Gyroscopic Unit. *Izvestiia vuzov. Priborostroenie*, 2015, vol. 58, no. 10, pp. 157–166 (In Russian).
11. Burakov M. V., Konovalov A. S., Yakovets O. B. Evolutionary Synthesis of Fuzzy Controllers. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 6, pp. 28–33 (In Russian). doi:10.15217/issn1684-8853.2015.6.28
12. Pacejka H. B., Besselink I. Y. Magic Formula Tyre Model with Transient Properties. *Supplement to Vehicle System Dynamics*, 1997, vol. 27, pp. 234–249.
13. Oudghiri M., Chadli M., El Hajjaji A. Robust Fuzzy Sliding Mode Control for Antilock Braking System. *Intern. Journal on Sciences and Techniques of Automatic Control*, 2007, no. 1, pp. 13–28.
14. Chankit J., Rahul A., Abhishek D. Linear Control Technique for Anti-Lock Braking System. *Intern. Journal of Engineering Research and Applications*, 2014, vol. 4, no. 8, pp. 104–108.
15. Chen C. K., Shin M. C. PID-Type Fuzzy Control for Anti-Lock Brake Systems with Parameter Adaptation. *JSME Intern. Journal. Ser. C*, 2004, vol. 47, no. 2, pp. 675–685.

СИСТЕМА ПЕРЕДАЧИ ИНФОРМАЦИИ, ОСНОВАННАЯ НА ОБОБЩЕННОЙ ХАОТИЧЕСКОЙ СИНХРОНИЗАЦИИ

Д. Д. Кульминский^а, младший научный сотрудник

В. И. Пономаренко^а, доктор физ.-мат. наук, доцент

М. Д. Прохоров^а, доктор физ.-мат. наук, доцент

Б. П. Безручко^б, доктор физ.-мат. наук, профессор

^аСаратовский филиал Института радиотехники и электроники им. В. А. Котельникова РАН, Саратов, РФ

^бСаратовский государственный университет им. Н. Г. Чернышевского, Саратов, РФ

Введение: системы передачи информации, основанные на режиме обобщенной хаотической синхронизации между передатчиком и приемником, традиционно используют два идентичных генератора в приемнике, создание которых зачастую является технически сложной задачей. **Цель:** разработка и исследование схемы передачи информации на основе обобщенной хаотической синхронизации, приемник которой содержит лишь один генератор, что снимает вопрос об обеспечении идентичности двух генераторов в приемнике. **Результаты:** предложена система передачи информации, основанная на использовании режима обобщенной хаотической синхронизации, которая содержит лишь одну ведомую автоколебательную систему в приемнике. Для диагностики режима обобщенной синхронизации между ведущей системой (передатчиком) и ведомой системой (приемником) предложено подавать на единственную ведомую систему поочередно сигнал ведущей системы и его задержанную копию и проверять, будет ли ведомая система демонстрировать одинаковую динамику. В отсутствие обобщенной синхронизации динамика ведомой системы оказывается разной при воздействии на нее два раза одним и тем же сигналом, а при наличии обобщенной синхронизации ведомая система демонстрирует после переходного процесса одинаковые колебания в обоих случаях. Работоспособность предложенной системы передачи информации продемонстрирована для случая, когда в качестве ведущей и ведомой систем использованы генераторы с запаздывающей обратной связью. Проведены численные исследования предложенной схемы связи. Выявлено, что она обладает высокой устойчивостью к шумам в канале связи. Разработанная система передачи информации реализована в радиофизическом эксперименте. Показана ее эффективность при передаче бинарного информационного сигнала. **Практическая значимость:** предложенная схема связи позволяет избежать технических трудностей создания систем передачи информации на основе обобщенной синхронизации, связанных с необходимостью обеспечить в эксперименте идентичность двух генераторов в приемнике.

Ключевые слова — система передачи информации, обобщенная синхронизация, системы с запаздыванием.

Введение

Системы связи, принцип действия которых базируется на использовании такого нелинейного явления, как синхронизация однонаправленно связанных хаотических систем, уже давно активно разрабатываются многими авторами [1–10]. В большинстве предложенных схем передачи информации, построенных на хаотической синхронизации, используется режим полной синхронизации между генераторами передатчика и приемника, представляющими собой ведущую и ведомую системы соответственно. Известны также системы передачи информации, основанные на фазовой синхронизации хаотических генераторов [11]. Следует отметить, что для систем связи как с полной, так и с фазовой синхронизацией очень важна высокая степень идентичности параметров приемника и передатчика, которую трудно обеспечить в реальных экспериментальных системах. Кроме того, такие информационно-коммуникационные системы критичны к шумам в канале связи и, как правило, теряют работоспособность уже при сравнительно небольших уровнях

шума, при которых отношение сигнал/шум оказывается менее 30 дБ [12].

Существенно повысить устойчивость к шуму и флуктуациям в канале связи удастся в системах передачи информации, основанных на явлении обобщенной хаотической синхронизации [13–15]. Для диагностики режима обобщенной синхронизации между ведущей системой (передатчиком) и ведомой системой, находящейся в приемнике, традиционно используется метод, при котором в приемник дополнительно вводится вспомогательная система [16], являющаяся точной копией ведомой системы. Хотя при этом не требуется идентичность генераторов в приемнике и передатчике, возникает задача построения двух идентичных генераторов в приемнике. Ее решение зачастую является технически сложным, особенно при работе на высоких частотах.

В данной работе мы предлагаем систему передачи информации, основанную на обобщенной хаотической синхронизации, в которой не требуется вспомогательная система в приемнике, за счет чего снимается вопрос об обеспечении идентичности используемых генераторов. Предложенная

схема связи исследована численно и реализована в радиофизическом эксперименте. Показано, что схема демонстрирует высокую устойчивость к шуму.

Схема передачи информации

Блок-схема предложенной нами системы связи на основе обобщенной синхронизации представлена на рис. 1.

Передатчик содержит хаотическую ведущую систему $x(t)$, на которую действует бинарный информационный сигнал $m(t)$, модулируя один из ее параметров. Сигнал с выхода передатчика передается в канал связи. В состав приемника входят автоколебательная ведомая система $y(t)$, две одинаковые линии задержки со временем запаздывания τ , коммутатор, генератор прямоугольных импульсов и разностный усилитель. Параметры схемы выбраны так, чтобы при передаче бинарного нуля между ведущей и ведомой системами существовал режим обобщенной синхронизации, а при передаче бинарной единицы обобщенной синхронизации не было.

Для диагностики существования обобщенной синхронизации между ведущей и ведомой системами мы не используем в приемнике вспомогательную систему, как это обычно делается в системах связи, основанных на обобщенной синхронизации [13–15]. Вместо этого мы подаем на единственную ведомую систему поочередно сигнал ведущей системы и его задержанную копию. Если обобщенная синхронизация между ведущей и ведомой системами существует, то после переходного процесса ведомая система будет демонстрировать идентичные колебания при воздействии на нее одинаковым сигналом. Если же обобщенная синхронизация между ведущей и ведомой системами отсутствует, то колебания ведомой системы будут разными, несмотря на то, что на нее действует один и тот же сигнал.

Для того чтобы два раза подействовать на ведомую систему одинаковым сигналом, мы исполь-

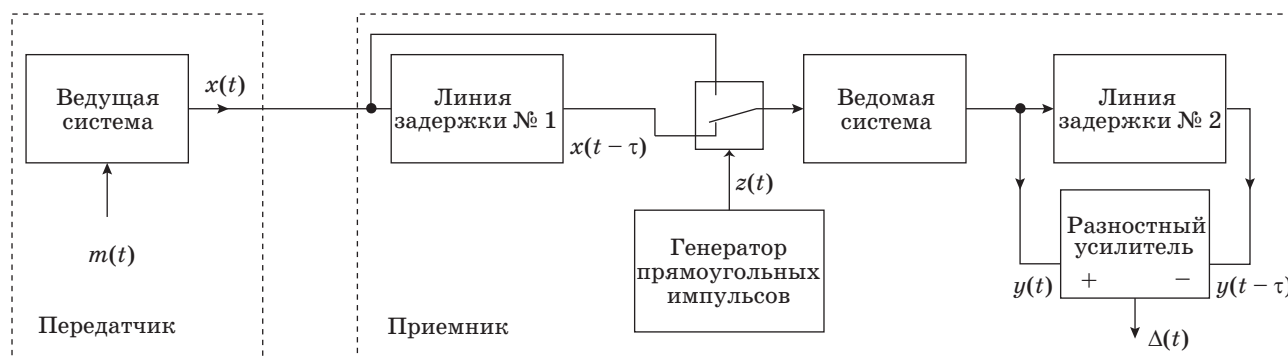
зуем линию задержки № 1 (см. рис. 1). Переключения сигнала воздействия на ведомую систему осуществляет коммутирующее устройство, которое управляется сигналом генератора прямоугольных импульсов $z(t)$, имеющим период T . Половину периода сигнала $z(t)$ ведомая система находится под действием сигнала $x(t)$, а другую половину периода на ведомую систему действует сигнал $x(t - \tau)$. Для работоспособности схемы длительность переходного процесса, предшествующего возникновению режима обобщенной синхронизации, должна быть меньше $T/2$. Время задержки следует выбрать равным $\tau = T/2$.

Чтобы сравнить колебания ведомой системы при воздействии на нее сигналов $x(t)$ и $x(t - \tau)$, мы использовали в приемнике линию задержки № 2 со временем запаздывания τ и разностный усилитель. При наличии обобщенной синхронизации между ведущей и ведомой системами разность сигналов $\Delta(t) = y(t) - y(t - \tau)$ на выходе разностного усилителя будет равна нулю во второй половине периода $z(t)$ после окончания переходного процесса. При отсутствии обобщенной синхронизации разность $\Delta(t)$ будет отлична от нуля в течение всей второй половины периода сигнала $z(t)$.

На протяжении всей первой половины периода $z(t)$ наблюдаются незатухающие колебания сигнала $\Delta(t)$, которые внешне похожи на колебания этого сигнала во второй половине периода $z(t)$ в отсутствие обобщенной синхронизации между ведущей и ведомой системами. Таким образом, в течение этого интервала времени разность $\Delta(t)$ не несет полезной информации для диагностики режима обобщенной синхронизации.

Численное исследование системы передачи информации

В качестве ведущей и ведомой систем в предложенной схеме передачи информации могут быть использованы различные генераторы хаотических колебаний. Мы рассмотрим случай, когда



■ Рис. 1. Блок-схема системы передачи информации

в качестве хаотических генераторов в передатчике и приемнике используются генераторы с запаздывающей обратной связью.

Ведущая система состоит из двух линий задержки (№ 3 со временем запаздывания τ_1 и № 4 со временем запаздывания τ_2), коммутатора, нелинейного элемента № 1 и линейного фильтра нижних частот № 1 (рис. 2, а). Коммутатор управляется бинарным информационным сигналом $m(t)$ и переключает время запаздывания в системе так, что при передаче бинарного нуля оно равно τ_1 , а при передаче бинарной единицы равно $\tau_1 + \tau_2$. Такая система описывается дифференциальным уравнением с запаздыванием

$$\varepsilon_1 \dot{x}(t) = -x(t) + f_1 \left(x(t - (\tau_1 + m(t)\tau_2)) \right), \quad (1)$$

где ε_1 — параметр инерционности, а f_1 — нелинейная функция.

Ведомая система состоит из линии задержки № 5 со временем запаздывания τ_3 , нелинейного элемента № 2, сумматора и линейного фильтра нижних частот № 2 (рис. 2, б). На нее поочередно действуют сигналы $x(t)$ и $x(t - \tau)$. Ведомая система описывается уравнением

$$\varepsilon_2 \dot{y}(t) = -y(t) + f_2 \left(y(t - \tau_3) \right) + k \left(z(t)x(t) + \overline{z(t)}x(t - \tau) \right), \quad (2)$$

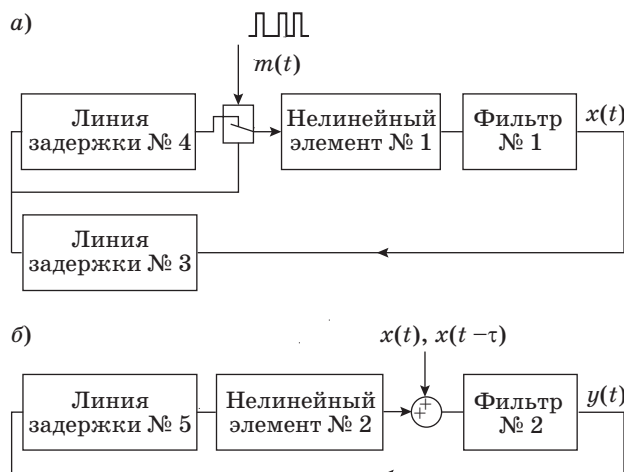
где ε_2 — параметр инерционности; f_2 — нелинейная функция; k — коэффициент связи; $z(t)$ — инверсия сигнала $x(t)$. При воздействии на ведомую систему сигналом $x(t)$ имеем $z(t) = 1$ и $\overline{z(t)} = 0$, а при воздействии сигналом $x(t - \tau)$ имеем $z(t) = 0$ и $\overline{z(t)} = 1$.

Пусть нелинейные элементы № 1 и № 2 имеют квадратичную нелинейность ($f_1(x) = \lambda_1 - x^2$ и $f_2(y) = \lambda_2 - y^2$ соответственно, где λ_1 и λ_2 — параметры нелинейности), а фильтры № 1 и № 2

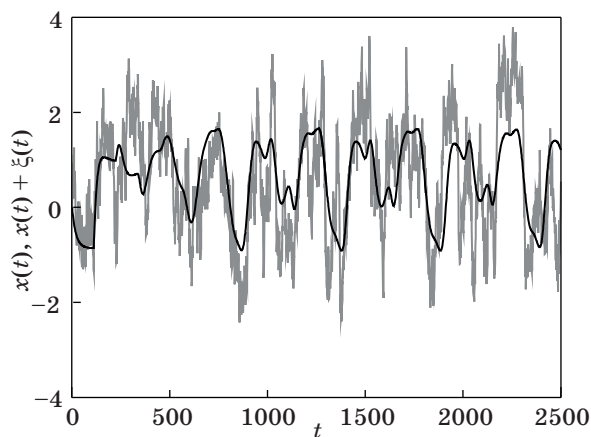
являются фильтрами Баттерворта первого порядка и имеют частоту среза $\nu_1 = 1/\varepsilon_1$ и $\nu_2 = 1/\varepsilon_2$ соответственно. Параметры системы передачи информации выберем следующими: $\tau_1 = 110$, $\tau_2 = 10$, $\tau_3 = 100$, $\lambda_1 = 1,7$, $\lambda_2 = 1,3$, $\varepsilon_1 = 20$, $\varepsilon_2 = 25$, $k = 0,08$, $\tau = 20\,000$. При этом передатчик генерирует хаотический сигнал (рис. 3), а в приемнике при $k = 0$ наблюдаются периодические колебания. В системах связи, основанных на обобщенной синхронизации, использование генератора периодических колебаний в приемнике имеет преимущества перед использованием генератора хаотических колебаний [17]. Обратим внимание, что передатчик и приемник имеют также расстройку по времени запаздывания и параметру инерционности, чтобы исключить возникновение режима полной синхронизации между ними.

Мы исследовали устойчивость предложенной схемы передачи информации к шуму, добавляя к временному ряду хаотического сигнала $x(t)$, передаваемого в канал связи, шум $\xi(t)$ различной интенсивности, имеющий нормальное распределение, нулевое среднее и отфильтрованный таким же фильтром, как фильтр № 1 ведущей системы. Временная реализация зашумленного сигнала для случая, когда отношение сигнал/шум составляло -6 дБ, приведена на рис. 3.

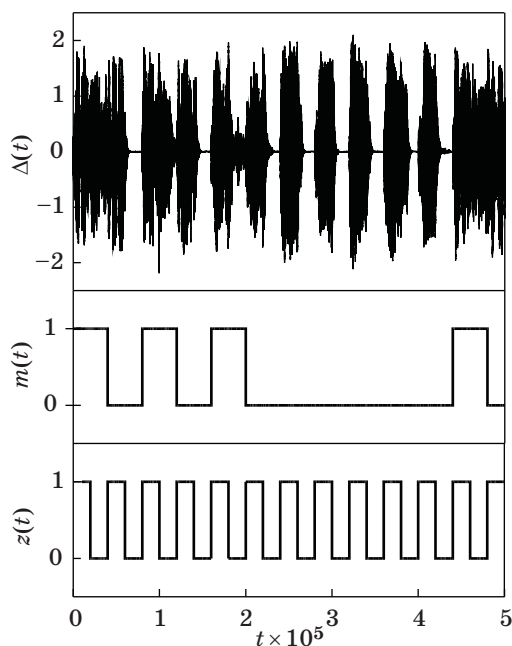
На рис. 4 показаны временные реализации информационного сигнала $m(t)$, управляющего сигнала $z(t)$ и сигнала $\Delta(t) = y(t) - y(t - \tau)$ на выходе разностного усилителя при отношении сигнал/шум, равном -6 дБ. При $m(t) = 0$ разность $\Delta(t)$ демонстрирует незатухающие колебания при $z(t) = 1$ и стремится к нулю при $z(t) = 0$, что указывает на существование обобщенной синхронизации между $x(t)$ и $y(t)$. При передаче бинарной единицы ($m(t) = 1$) сигнал $\Delta(t)$ отличен от нуля при любом значении $z(t)$, что свидетельствует



■ Рис. 2. Блок-схемы ведущей (а) и ведомой (б) систем



■ Рис. 3. Фрагменты временных реализаций хаотического сигнала $x(t)$ в отсутствие шума (черная кривая) и при наличии интенсивного аддитивного шума (серая кривая)



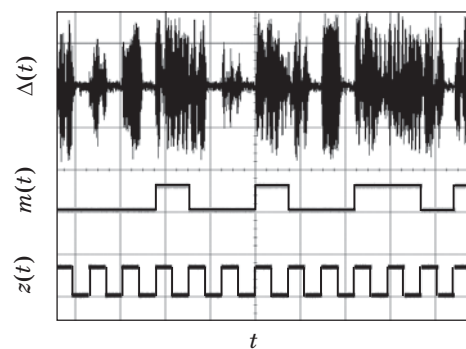
■ Рис. 4. Фрагменты временных реализаций сигналов $\Delta(t)$, $m(t)$ и $z(t)$

об отсутствии обобщенной синхронизации между ведущей и ведомой системами. Несмотря на очень высокий уровень шума, предложенная система связи позволяет выделить бинарный информационный сигнал.

Экспериментальное исследование системы передачи информации

Предложенная система связи на основе обобщенной синхронизации реализована нами в радиофизическом эксперименте. Ведущая и ведомая системы в экспериментальной схеме представляли собой гибридные генераторы с запаздывающей обратной связью, состоящие из цифровых и аналоговых элементов. Линии задержки и нелинейные элементы были реализованы в цифровом виде на базе программируемых микроконтроллеров, а фильтры представляли собой аналоговые низкочастотные RC -фильтры первого порядка. Вспомогательные линии задержки № 1 и № 2 в приемнике (см. рис. 1) тоже были реализованы в цифровом виде с помощью микроконтроллера.

Параметры ведущей системы, описываемой уравнением (1), имели следующие значения: $\tau_1 = 1,244$ мс, $\tau_2 = 113$ мкс, $\varepsilon_1 = R_1 C_1 = 113$ мкс, $f_1(x) = \lambda_1 - x^2$, где $\lambda_1 = 1,7$, — соответствующие хаотическому режиму колебаний. Параметры ведомой системы, описываемой уравнением (2), были выбраны следующими: $\tau_3 = 1,13$ мс, $\varepsilon_2 = R_2 C_2 = 287$ мкс, $f_2(y) = \lambda_2 - y^2$, где $\lambda_2 = 1,3$, $k = 0,138$, $\tau = 185$ мс. При таких параметрах ве-



■ Рис. 5. Осциллограммы временных реализаций фильтрованного разностного сигнала $\Delta(t)$, информационного сигнала $m(t)$ и сигнала $z(t)$ генератора прямоугольных импульсов

мая система в отсутствие связи находилась в режиме периодических колебаний.

Фрагменты экспериментальных временных реализаций разностного сигнала $\Delta(t) = y(t) - y(t - \tau)$, информационного сигнала $m(t)$ и управляющего сигнала $z(t)$ показаны на рис. 5. Для лучшей визуализации режима обобщенной синхронизации сигнал $\Delta(t)$ отфильтрован фильтром нижних частот с частотой среза $\nu = 200$ Гц. Временной масштаб по горизонтальной оси равен 500 мс/дел, а масштаб по вертикальной оси равен 200 мВ/дел для $\Delta(t)$ и 5 В/дел для $m(t)$ и $z(t)$.

Как видно из рис. 5, при передаче бинарной единицы амплитуда колебаний сигнала $\Delta(t)$, так же как и в рассмотренном выше численном примере, практически не зависит от амплитуды сигнала $z(t)$, что свидетельствует об отсутствии обобщенной синхронизации между ведущей и ведомой системами. При передаче бинарного нуля амплитуда колебаний сигнала $\Delta(t)$, хотя и не падает до нуля, оказывается заметно меньше при более низких значениях $z(t)$. Такое резкое уменьшение амплитуды $\Delta(t)$ при изменении $z(t)$ является признаком того, что обобщенная синхронизация между ведущей и ведомой системами существует.

Заключение

Нами предложена система передачи информации, основанная на использовании режима обобщенной синхронизации между передатчиком и приемником. В отличие от других систем связи на основе обобщенной синхронизации, предложенная схема содержит лишь одну ведомую автоколебательную систему в приемнике. Отсутствие вспомогательной системы в приемнике позволяет избежать технических трудностей, характерных для основанных на обобщенной синхронизации систем связи и обусловленных необходимостью

обеспечить в эксперименте идентичность двух генераторов в приемнике.

Для диагностики режима обобщенной синхронизации между ведущей и ведомой системами предложено поочередно воздействовать на ведомую систему сигналом ведущей системы и задержанной копией этого сигнала. Работоспособность рассмотренной системы передачи информации

продемонстрирована для случая, когда в качестве ведущей и ведомой систем использованы генераторы с запаздывающей обратной связью.

Предложенная схема связи исследована численно и реализована экспериментально. Показана ее высокая устойчивость к шуму в канале связи.

Работа выполнена при поддержке РФФИ, грант № 16-02-00091.

Литература

1. Kocarev L., et al. Experimental Demonstration of Secure Communications via Chaotic Synchronization / L. Kocarev, K. S. Halle, K. Eckert, L. O. Chua, U. Parlitz // International Journal of Bifurcation and Chaos. 1992. Vol. 2. P. 709–713. doi:10.1142/S0218127492000823
2. Parlitz U., et al. Transmission of Digital Signals by Chaotic Synchronization / U. Parlitz, L. O. Chua, L. Kocarev, K. S. Halle, A. Shang // International Journal of Bifurcation and Chaos. 1992. Vol. 2. P. 973–977. doi:10.1142/S0218127492000562
3. Cuomo K. M., Oppenheim A. V. Circuit Implementation of Synchronized Chaos with Applications to Communications // Physical Review Letters. 1993. Vol. 71. P. 65–68. doi:10.1103/PhysRevLett.71.65
4. García-Ojalvo J., Roy R. Spatiotemporal Communication with Synchronized Optical Chaos // Physical Review Letters. 2001. Vol. 86. P. 5204–5207. doi:10.1103/PhysRevLett.86.5204
5. Дмитриев А. С., Панас А. И. Динамический хаос: новые носители информации для систем связи. — М.: Физматлит, 2002. — 252 с.
6. Ponomarenko V. I., Prokhorov M. D. Extracting Information Masked by the Chaotic Signal of a Time-Delay System // Physical Review E. 2002. Vol. 66. 026215. doi:10.1103/PhysRevE.66.026215
7. Караваев А. С., Кульминский Д. Д., Пономаренко В. И., Прохоров М. Д. Система цифровой передачи информации, маскируемой хаотическим сигналом системы с запаздыванием // Информационно-управляющие системы. 2013. № 4(65). С. 30–35.
8. Abderrahim N. W., Benmansour F. Z., Seddiki O. A Chaotic Stream Cipher Based on Symbolic Dynamic Description and Synchronization // Nonlinear Dynamics. 2014. Vol. 78. P. 197–207. doi:10.1007/s11071-014-1432-z
9. Кульминский Д. Д., Пономаренко В. И., Караваев А. С., Прохоров М. Д. Система связи, основанная на синхронизации систем с задержкой с переключением хаотических режимов // Информационно-управляющие системы. 2015. № 3(76). С. 85–91. doi:10.15217/issn1684-8853.2015.3.85
10. Karavaev A. S., Kulminskiy D. D., Ponomarenko V. I., Prokhorov M. D. An Experimental Communication Scheme Based on Chaotic Time-Delay System with Switched Delay // International Journal of Bifurcation and Chaos. 2015. Vol. 25. 1550134. doi:10.1142/S0218127415501345
11. Chen J. Y., Wong K. W., Cheng L. M., Shuai J. W. A Secure Communication Scheme Based on the Phase Synchronization of Chaotic Systems // Chaos. 2003. Vol. 13. P. 508–514. doi: 10.1063/1.1564934
12. Короновский А. А., Москаленко О. И., Храмов А. Е. О применении хаотической синхронизации для скрытой передачи информации // Успехи физических наук. 2009. Т. 179. С. 1281–1310. doi:10.3367/UFNr.0179.200912c.1281
13. Короновский А. А., Москаленко О. И., Попов П. В., Храмов А. Е. Способ скрытой передачи информации, основанный на явлении обобщенной синхронизации // Известия РАН. Серия физическая. 2008. Т. 72. № 1. С. 143–147.
14. Moskalenko O. I., Koronovskii A. A., Hramov A. E. Generalized Synchronization of Chaos for Secure Communication: Remarkable Stability to Noise // Physics Letters A. 2010. Vol. 374. P. 2925–2931. doi:10.1016/j.physleta.2010.05.024
15. Короновский А. А., Москаленко О. И., Храмов А. Е. Скрытая передача информации на основе режима обобщенной синхронизации в присутствии шумов // Журнал технической физики. 2010. Т. 80. В. 4. С. 1–8.
16. Abarbanel H. D. I., Rulkov N. F., Sushchik M. M. Generalized Synchronization of Chaos: The Auxiliary System Approach // Physical Review E. 1996. Vol. 53. P. 4528–4535. doi:10.1103/PhysRevE.53.4528
17. Короновский А. А. и др. Обобщенная синхронизация в случае воздействия хаотического сигнала на периодическую систему / А. А. Короновский, О. И. Москаленко, А. С. Павлов, Н. С. Фролов, А. Е. Храмов // Журнал технической физики. 2014. Т. 84. Вып. 5. С. 1–8.

UDC 537.86

doi:10.15217/issn1684-8853.2016.2.42

Communication System based on Generalized Chaotic SynchronizationKul'minskii D. D.^a, Junior Researcher, kulminskydd@gmail.comPonomarenko V. I.^a, Dr. Sc., Phys.-Math., Associate Professor, ponomarenkovi@gmail.comProkhorov M. D.^a, Dr. Sc., Phys.-Math., Associate Professor, mdprokhorov@yandex.ruBezruchko B. P.^b, Dr. Sc., Phys.-Math., Professor, bezruchkobp@gmail.com^aSaratov Branch of Kotel'nikov Institute of Radio Engineering and Electronics of RAS, 38, Zelyonaya St., 410019, Saratov, Russian Federation^bSaratov State University, 83, Astrakhanskaya St., 410012, Saratov, Russian Federation

Introduction: Communication systems based on the regime of generalized chaotic synchronization between transmitter and receiver conventionally use two identical generators in the receiver. Their development is often a technically difficult problem. **Purpose:** The goal is the development and research of a communication scheme based on generalized chaotic synchronization when the receiver contains only one generator. This would eliminate the problem of ensuring the identity of two generators in the receiver. **Results:** A data transmission scheme is proposed, based on the regime of generalized chaotic synchronization, which has only one self-oscillating response system in the receiver. To detect the regime of generalized synchronization between the drive system (transmitter) and response system (receiver), a single response system is affected in turn by the drive system signal and its delayed copy. After that, it is checked whether the response system exhibits similar dynamics. In the absence of generalized synchronization, the response system shows different oscillations under the driving by the same signal, but in the presence of generalized synchronization, the response system dynamics is identical in both cases after the transient process is over. The efficacy of the proposed communication system is shown for the case when time-delay generators are used as the drive and response systems. Numerical investigation of the scheme have been carried out. The scheme has shown a high resistance to noise in the transmission channel. The developed communication system is experimentally implemented. Its efficacy has been demonstrated for the case of binary information signal transmission. **Practical relevance:** The proposed system of information transmission allows one to avoid the technical difficulties common in communication systems based on generalized synchronization caused by the necessity to create two identical generators in the receiver.

Keywords — Communication System, Generalized Synchronization, Time-Delay Systems.

References

- Kocarev L., Halle K. S., Eckert K., Chua L. O., Parlitz U. Experimental Demonstration of Secure Communications via Chaotic Synchronization. *International Journal of Bifurcation and Chaos*, 1992, vol. 2, pp. 709–713. doi:10.1142/S0218127492000823
- Parlitz U., Chua L. O., Kocarev L., Halle K. S., Shang A. Transmission of Digital Signals by Chaotic Synchronization. *International Journal of Bifurcation and Chaos*, 1992, vol. 2, pp. 973–977. doi:10.1142/S0218127492000562
- Cuomo K. M., Oppenheim A. V. Circuit Implementation of Synchronized Chaos with Applications to Communications. *Physical Review Letters*, 1993, vol. 71, pp. 65–68. doi:10.1103/PhysRevLett.71.65
- García-Ojalvo J., Roy R. Spatiotemporal Communication with Synchronized Optical Chaos. *Physical Review Letters*, 2001, vol. 86, pp. 5204–5207. doi:10.1103/PhysRevLett.86.5204
- Dmitriev A. S., Panas A. I. *Dinamicheskie khaos: novye nositeli informatsii dlya sistem svyazi* [Dynamical Chaos: New Information Carriers for Communication Systems]. Moscow, Fizmatlit Publ., 2002. 252 p. (In Russian).
- Ponomarenko V. I., Prokhorov M. D. Extracting Information Masked by the Chaotic Signal of a Time-Delay System. *Physical Review E*, 2002, vol. 66, 026215. doi:10.1103/PhysRevE.66.026215
- Karavaev A. S., Kul'minskii D. D., Ponomarenko V. I., Prokhorov M. D. System of Digital Transmission of Information Masked by Chaotic Signal of Time-Delay System. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2013, no. 4(65), pp. 30–35 (In Russian).
- Abderrahim N. W., Benmansour F. Z., Seddiki O. A Chaotic Stream Cipher Based on Symbolic Dynamic Description and Synchronization. *Nonlinear Dynamics*, 2014, vol. 78, pp. 197–207. doi:10.1007/s11071-014-1432-z
- Kul'minskii D. D., Ponomarenko V. I., Karavaev A. S., Prokhorov M. D. Communication System Based on Synchronization of Time-Delay Systems with Switching of Chaotic Regimes. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2015, no. 3(76), pp. 85–91 (In Russian). doi:10.15217/issn1684-8853.2015.3.85
- Karavaev A. S., Kulminskiy D. D., Ponomarenko V. I., Prokhorov M. D. An Experimental Communication Scheme Based on Chaotic Time-Delay System with Switched Delay. *International Journal of Bifurcation and Chaos*, 2015, vol. 25, 1550134. doi:10.1142/S0218127415501345
- Chen J. Y., Wong K. W., Cheng L. M., Shuai J. W. A Secure Communication Scheme Based on the Phase Synchronization of Chaotic Systems. *Chaos*, 2003, vol. 13, pp. 508–514. doi: 10.1063/1.1564934
- Koronovskii A. A., Moskalenko O. I., Hramov A. E. On the Use of Chaotic Synchronization for Secure Communication. *Uspekhi fizicheskikh nauk*, 2009, vol. 179, pp. 1213–1238 (In Russian). doi:10.3367/UFNe.0179.200912c.1281
- Koronovskii A. A., Moskalenko O. I., Popov P. V., Hramov A. E. Method for Secure Data Transmission Based on Generalized Synchronization. *Izvestiya RAN. Seriya fizicheskaya*, 2008, vol. 72, pp. 131–135 (In Russian). doi:10.3103/S1062873808010309
- Moskalenko O. I., Koronovskii A. A., Hramov A. E. Generalized Synchronization of Chaos for Secure Communication: Remarkable Stability to Noise. *Physics Letters A*, 2010, vol. 374, pp. 2925–2931. doi:10.1016/j.physleta.2010.05.024
- Koronovskii A. A., Moskalenko O. I., Hramov A. E. Hidden Data Transmission Using Generalized Synchronization in the Presence of Noise. *Zhurnal tekhnicheskoi fiziki*, 2010, vol. 55, pp. 435–441 (In Russian). doi:10.1134/S1063784210040018
- Abarbanel H. D. I., Rulkov N. F., Sushchik M. M. Generalized Synchronization of Chaos: The Auxiliary System Approach. *Physical Review E*, 1996, vol. 53, pp. 4528–4535. doi:10.1103/PhysRevE.53.4528
- Koronovskii A. A., Moskalenko O. I., Pavlov A. S., Frolov N. S., Hramov A. E. Generalized Synchronization in the Action of a Chaotic Signal on a Periodic System. *Zhurnal tekhnicheskoi fiziki*, 2014, vol. 84, iss. 5, pp. 1–8 (In Russian). doi:10.1134/S1063784214050132

К ОЦЕНИВАНИЮ И СОПОСТАВЛЕНИЮ ОЧЕРЕДЕЙ КЛАССИЧЕСКИХ И ФРАКТАЛЬНЫХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

О. И. Кутузов^а, доктор техн. наук, профессор

Т. М. Татарникова^б, доктор техн. наук, доцент

^аСанкт-Петербургский государственный электротехнический университет «ЛЭТИ», Санкт-Петербург, РФ

^бСанкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ

Постановка проблемы: буферные накопители сетевых узлов являются важнейшим ресурсом управления трафиком. Исследования последних полутора десятков лет доказывают, что сетевой трафик по своей природе является самоподобным, и поэтому традиционный анализ очередей, основанный на предположении о пуассоновском потоке, не позволяет точно предсказать производительность узла и сети в целом. Отсутствие общих аналитических результатов исследования очередей при самоподобном трафике диктует необходимость искать пути его адекватного описания и моделирования. **Цель:** разработка модели поступлений в самоподобном трафике, позволяющей оценить и сопоставить в первом приближении характеристики очередей классических и фрактальных систем массового обслуживания. **Результаты:** предложена модель поступлений в самоподобном трафике, отличающаяся присутствием долговременных зависимостей в трассах трафика и наличием пульсирующей структуры на многих масштабах времени, что позволяет адекватно отразить свойства сетевого трафика. Предложены модели расчета буферных накопителей сетевых узлов, являющихся более реалистичными в условиях самоподобного трафика. **Практическая значимость:** модель поступлений в самоподобном трафике и модели расчета буферных накопителей могут найти применение при решении задач управления сетевым трафиком и планировании ресурсов сети.

Ключевые слова — самоподобный трафик, пульсирующая структура трафика, ON/OFF-модель, долговременная зависимость, распределения с тяжелыми «хвостами», распределение Парето, системы с очередями, буферный накопитель, вероятность потерь, формула Полячека — Хинчина, конечная очередь, бесконечная очередь, управление сетевым трафиком.

Введение

Буферные накопители сетевых узлов являются важнейшим ресурсом управления сетевым трафиком. Исследования последних полутора десятков лет доказывают, что сетевой трафик по своей природе является самоподобным (self-similar) или фрактальным (fractal) [1, 2]. К фрактальным свойствам, характеризующим трафик, обычно относят такие понятия, как «медленно затухающие дисперсии», «долговременная зависимость» [3].

Особенностью самоподобного трафика является устойчивость кластеризации [4]. Традиционные модели пакетного трафика являются кратковременно зависимыми, т. е. имеют экспоненциально затухающие корреляции. Данные самоподобного трафика проявляют долговременную зависимость и имеют гиперболически затухающие корреляции. Самоподобность в математическом смысле является геометрическим понятием. Оно подчеркивает то, что модели, описывающие разные типы трафика, могут иметь одинаковую структуру, задаваемую разномасштабными пульсациями. Понятие пульсирующей структуры трафика часто используется именно в этом контексте [4, 5].

Самоподобность трафика оказывает существенное влияние на качество связи. Исследования в большинстве случаев концентрируются вокруг статистических характеристик очередей, поскольку буферизация является ведущей обеспечивающей ресурсами стратегией. Оказывается, что традиционный анализ очередей, базирующийся на предположении о пуассоновском потоке, не может точно предсказать производительность системы в условиях самоподобного трафика: «...практическое следствие фрактальной структуры трафика: буферы коммутаторов и мультиплексоров должны иметь значительно больший объем, чем это предписывалось результатами традиционного анализа и моделирования процессов обработки очередей» [4].

Адекватное описание самоподобного трафика определяется распределениями вероятностей с тяжелыми «хвостами». В настоящее время не существует общих аналитических результатов исследования очередей при самоподобном трафике и влияния самоподобности на качество его обслуживания [6].

Ограниченные возможности применения точных и приближенных аналитических методов приводят к необходимости проводить имитационное моделирование, которое, однако, сопря-

жено с рядом специфических вычислительных проблем, обусловленных стохастическим характером отклика имитационных моделей и необходимостью корректной обработки этого отклика [7, 8].

Тем не менее представляется целесообразным в первом приближении оценить и сопоставить характеристики очередей классических и фрактальных систем массового обслуживания (СМО).

Модель поступлений в самоподобном трафике

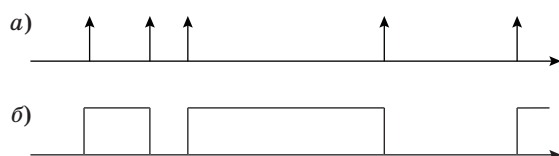
При моделировании самоподобного трафика необходимо отразить присутствие долговременных зависимостей в трассах трафика и его пульсирующую (самоподобную) структуру на многих масштабах времени. Оба эти эффекта могут моделироваться пиковым процессом. Пик подразумевает длительный поток блоков данных, например пакетов или ячеек, которые вбрасываются в сеть отдельным пользователем, допустим, при передаче данных по FTP-протоколу. Длина пика при этом описывается распределениями с тяжелыми «хвостами» [9].

Для описания компьютерного сетевого трафика широко используется чередующийся фрактальный возобновляющийся процесс, известный как ON/OFF-модель, или модель серии пакетов (рис. 1, а и б).

Известно, что любой источник поступлений определяется двумя состояниями: активным периодом (ON-периодом), когда источник передает пакеты, и пассивным — когда источник ничего не передает (OFF-периодом). Каждый переход от состояния OFF к состоянию ON соответствует поступающему пакету или серии пакетов. Продолжительности того или иного состояния являются независимыми одинаково распределенными и не обязательно одинаковыми для периодов ON и OFF. На практике при моделировании источников распределение OFF-периодов часто полагают экспоненциальным, а для ON-периодов наиболее часто используется распределение Парето [2, 5].

Распределение Парето определяется функцией распределения в виде

$$F(x) = 1 - \left(\frac{K}{x}\right)^\alpha, \quad x \geq K, \quad (1)$$



■ Рис. 1. Стандартный фрактальный возобновляющийся процесс (а) и фрактальный ON/OFF-процесс (б)

где K — граничный параметр, задающий минимальное значение случайной величины x , является, по сути, масштабным коэффициентом; α — параметр формы.

Математическое ожидание и дисперсия случайной величины, распределенной по Парето, определяются, соответственно, следующими выражениями:

$$M(x) = \frac{\alpha K}{\alpha - 1}; \quad (2)$$

$$D(x) = \frac{\alpha K^2}{(\alpha - 1)^2 (\alpha - 2)}. \quad (3)$$

Если $\alpha > 1$, то математическое ожидание конечно, дисперсия конечна только при $\alpha > 2$.

Распределение Парето имеет «хвост» $P(x > t) = 1 - F(t)$:

$$1 - F(t) = 1 - \left[1 - \left(\frac{K}{t}\right)^\alpha\right] = \left(\frac{K}{t}\right)^\alpha \propto t^{-\alpha}.$$

Как видно, при любых значениях K и α скорость убывания «хвоста» распределения Парето степенная. Далее распределение Парето будем обозначать в виде $Pa(K, \alpha)$.

Кроме математического ожидания и дисперсии, полезной характеристикой случайной величины x при моделировании очередей является квадрат ее коэффициента вариации C_x . Для $x \in Pa(K, \alpha)$ из (2) и (3) находим

$$C_x^2 = \frac{D(x)}{M^2(x)} = \frac{\alpha K^2}{(\alpha - 1)^2 (\alpha - 2)} \frac{(\alpha - 1)^2}{(\alpha K)^2} = \frac{1}{\alpha(\alpha - 2)}.$$

Коэффициенты вариации (точнее, их квадраты) являются мерой разброса значений для интервалов поступления заявок и интервалов их обслуживания. Чем эти коэффициенты больше, тем большего значения средней длины очереди $L(\rho)$ следует ожидать при фиксированной нагрузке ρ .

Для Парето-распределения в диапазоне значений параметра $1 < \alpha \leq 2$ оба квадрата коэффициентов вариации (интервалов между поступлениями и продолжительностей поступлений), по сути, стремятся к бесконечности, из чего следует, что и значения средней длины очереди $L(\rho)$ следует ожидать существенно (многократно) большего, чем для $M|M|1$.

Узел в сети как СМО

Для задач проектирования сетевых устройств одними из наиболее подходящих математических моделей являются системы с очередями (системы массового обслуживания) [10, 11].

В общем случае узел телекоммуникационной сети рассматриваем как СМО класса $G|G|1|m$

при стационарном режиме функционирования. Основную задачу проектирования буферов как систем класса $G|G|1|m$ поставим в виде нахождения зависимости вероятности потери заявки P от размера буфера m ($m = 0, 1, 2, \dots$):

$$P = \varphi(m). \quad (4)$$

Это позволит для любой наперед заданной допустимой вероятности потерь P^* определять соответствующий наименьший допустимый размер буфера m^* по формуле, следующей из (4):

$$\lfloor m^* \rfloor = \varphi^{-1}(P^*), \quad (5)$$

где φ^{-1} — функция, обратная функции φ ; $\lfloor x \rfloor$ — ближайшее целое не меньшее x .

Полагая, что функция φ монотонно убывающая, можно утверждать, что φ^{-1} существует.

Следовательно, решив основную задачу (4), для любой максимально допустимой вероятности потерь P^* можно по формуле (5) определять наименьший размер буфера m^* , обеспечивающий требуемый уровень качества сетевого сервиса.

Очевидно, модели с конечным буфером (КБ), $m < \infty$, ближе к реальным узлам телекоммуникационной сети, чем модели с неограниченным (бесконечным) буфером (ББ), $m = \infty$. Тем не менее модели с ББ также нередко используются для решения основной задачи (4). При этом предполагается, что СМО с КБ отличается от СМО с ББ лишь конечной емкостью буфера, а по всем остальным параметрам эти СМО идентичны. Однако в общем случае это не так. В СМО с КБ каждая заявка, приходящая в заполненную очередь длиной m , теряется и тем самым исключается из потока, что изменяет коэффициент загрузки СМО. В соответствующей СМО с ББ все заявки, приходящие в очередь, в которой m мест занято, остаются в ней, и еще некоторое время, пока не освободится обслуживающий прибор, длина очереди сохранится большей, чем m . Таким образом, длина очереди может достаточно долго превосходить m , что приводит к завышению искомой вероятности P [12].

Расчет буферной емкости классических СМО

В системе $M|M|1|m$ стационарная вероятность потери заявки $P(m)$ определяется выражением

$$P(m) = \rho^m \frac{1 - \rho}{1 - \rho^{m+1}}, \quad (6)$$

где ρ — загрузка системы, $\rho < 1$; m — величина допустимой очереди [10].

Выражение (6) получено из модели СМО с ББ. Задаваясь соответствующими значениями веро-

ятности $P(m)$ и разрешая (6) относительно требуемого значения m , получаем выражение

$$\lfloor m \rfloor = \left\lceil \ln \left(\frac{P(m)}{1 - \rho} \right) \ln \frac{1}{\rho} \right\rceil, \quad (7)$$

где $\lfloor m \rfloor$ — ближайшее целое не меньшее m .

Потери в КБ являются функцией от максимальной длины очереди в ББ на периоде регенерации [12]. Соответственно, вероятность потерь в КБ может быть рассчитана с использованием «хвоста» распределения вероятностей значений максимальной длины очереди в ББ. Распределение значений максимальной длины очереди зависит от характеристик сетевого трафика.

В работе [12] приведено выражение для оценивания вероятности превышения значения емкости m конечного буфера для системы $M|M|1|m$, полученное с использованием элементов теории максимальных статистик:

$$P(m) = R \frac{a_n}{n\bar{K}} \exp \left(-\frac{m - b_n + 1}{a_n} \right), \quad (8)$$

где R — нормирующий коэффициент дискретного распределения при использовании метода экстремальных статистик; n — число выборочных значений в последовательностях, на которые разбивается выборочная последовательность, составленная из максимальных значений очереди на интервалах регенерации; $\bar{K}$ — среднее число заявок, поступающих на вход СМО за период регенерации; a_n и b_n — коэффициенты, вычисляемые по выборочным данным.

Преобразуем выражение (8), разрешая его относительно значения m , для которого вероятность потери $P(m)$ не превышает допустимого значения $P_{\text{доп}}(m)$. Получим

$$m = a_n \ln R \frac{a_n}{n\bar{K}} - a_n \ln P_{\text{доп}}(m) + b_n - 1. \quad (9)$$

В табл. 1 представлены значения m , рассчитанные по формуле (7) для модели СМО типа $M|M|1$ с ББ и по формуле (9) для модели СМО типа $M|M|1$ с КБ. При расчетах по формуле (9)

■ Таблица 1

P_m	$M M 1$ с КБ			$M M 1$ с ББ		
	ρ					
	0,5	0,7	0,9	0,5	0,7	0,9
10^{-5}	14	23	82	16	29	92
10^{-7}	21	34	123	23	42	161
10^{-8}	25	39	144	26	48	184
10^{-9}	28	44	165	30	57	207
10^{-11}	35	55	206	36	67	230

использованы значения выборочных статистик $(\bar{K}, n, R, a_n, b_n)$ [12].

Результаты табл. 1 подтверждают ранее распространенное мнение о том, что расчет экспоненциальных СМО задает верхние границы оценок. Поэтому классические модели СМО широко использовались при расчете сетевых ресурсов.

Однако результаты нового взгляда на природу сетевого трафика послужили причиной проверить выводы, сделанные на основе расчетов с использованием традиционных моделей трафика.

Сопоставление очередей СМО $M|Pa|1$ и СМО $M|M|1$

У классических СМО зависимость $P(m)$ является асимптотически-экспоненциальной [13], т. е. с ростом размера буфера m вероятность потерь заявок P быстро падет и становится пренебрежимо малой. У фрактальных СМО с ростом m снижение P происходит очень медленно, и чтобы добиться требуемого малого ее значения, нужно предусматривать гораздо большие затраты оборудования на буферизацию, чем в случае классических СМО. Покажем это на примере сопоставления очередей у СМО типа $M|M|1$ и $M|Pa|1$, поскольку такое сопоставление можно провести как теоретически, используя формулу Полячека — Хинчина [14], так и экспериментально.

Для прямой процедуры обслуживания получено [15] отношение средних длин очередей $L_{с.п}$ и $L_{с.э}$ при Парето- и экспоненциальном распределении времени обслуживания соответственно и входящем пуассоновском потоке в виде

$$\frac{L_{с.п}}{L_{с.э}} = \frac{1}{2} \left[\frac{\alpha(\alpha-2)+1}{\alpha(\alpha-2)} \right]. \quad (10)$$

Производная выражения (10) отрицательна, т. е. с увеличением α значение отношения (10) устойчиво уменьшается и при $\alpha=1+\sqrt{2}$ становится равным 1. Следовательно, $\alpha=1+\sqrt{2}$ является пороговым значением, при превышении которого для экспоненциальной СМО $M|M|1$ средняя длина очереди оказывается большей, чем для СМО с пуассоновским входным потоком и распределенным по Парето временем обслуживания при одинаковой входящей нагрузке.

Пример значений отношения (10) при $\alpha > 2$ приведен в табл. 2 в сопоставлении с результатами численного эксперимента на имитационной модели одноканальной СМО. Также в табл. 2 приведены отношения $L_{м.п}/L_{м.э}$ максимальных значений очередей при распределении времени обслуживания по Парето- и экспоненциальном распределении, полученные экспериментальным путем. Объем выборки $N = 10^6$.

■ Таблица 2

Отношение очередей	α	ρ				Теоретический результат
		0,25	0,5	0,7	0,9	
$\frac{L_{с.п}}{L_{с.э}}$	3	0,63	0,63	0,62	0,62	0,66
	5	0,52	0,54	0,53	0,55	0,53
$\frac{L_{м.п}}{L_{м.э}}$	3	0,90	0,89	0,71	0,61	—
	5	0,58	0,61	0,56	0,54	—

Результаты эксперимента показывают, что при фрактальном трафике, длительности поступлений которого описываются распределением Парето со значениями параметра $\alpha \geq (1+\sqrt{2})$, средняя и максимальная длина очереди для экспоненциальной СМО $M|M|1$ превышают аналогичные характеристики очереди для СМО $M|Pa|1$ при одинаковой нагрузке для обеих СМО.

Следует отметить, что при анализе узлов телекоммуникационной сети в условиях фрактального трафика наиболее актуален диапазон значений параметра $\alpha \in [1, 2]$ [2]. При этих условиях формула Полячека — Хинчина не работает, соответственно, не работает и соотношение (10).

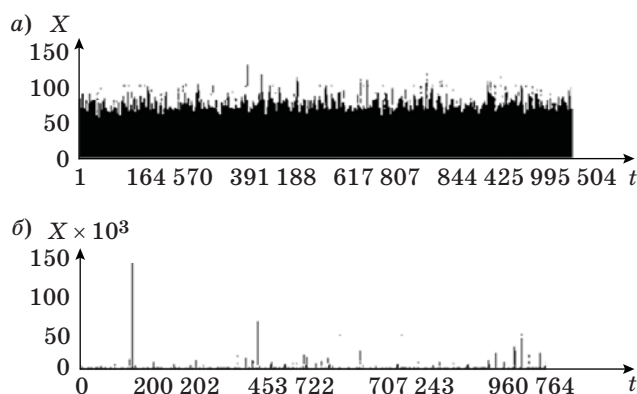
Оценивание влияния самоподобия на характеристики очереди при значениях $1 < \alpha \leq 2$ выполнено с применением численного эксперимента на имитационной модели соответствующей СМО [14]. Результаты эксперимента представлены в табл. 3.

Однако обратимся к выборкам из Парето- и экспоненциально распределенных значений случайных величин. Выборки выполнены без ограничения значений random (рис. 2).

Обращает на себя внимание тот факт, что максимальные значения ординат длительностей поступлений, распределенных по Парето, почти в 1000 раз превышают аналогичные значения при экспоненциальном распределении, хотя размеры поступлений отображают возможные длительности сообщений одного и того же трафика. Очевидно, что такой значительной разницы

■ Таблица 3

Отношение очередей	α	ρ			
		0,25	0,5	0,7	0,9
$\frac{L_{с.п}}{L_{с.э}}$	1,1	16 820,8	5708,0	1959,1	808,3
	1,5	12,7	12,9	14,6	24,1
	2,41	0,95	0,96	0,94	0,97
$\frac{L_{м.п}}{L_{м.э}}$	1,1	1384,9	1373,1	730,9	278,4
	1,5	51,5	44,5	35,0	22,9
	2,41	7,0	8,5	6,2	3,9



■ **Рис. 2.** Выборки поступлений: *a* — выборка при экспоненциальном распределении значений поступлений: $\lambda = 0,1$; *б* — выборка при Парето-распределении значений поступлений: $\alpha = 1,1, m = 10$

в отображении максимальных значений длительностей сообщений быть не должно.

Это различие изначально заложено в генерации реализаций случайной величины рассматриваемых распределений. Выборочные значения (длительностей сообщений) экспоненциально и Парето-распределенных случайных величин получаются преобразованием значений одного и того же датчика random. Однако максимальные выборочные значения этих распределений $y_{\max}^{\text{э}}$ и $y_{\max}^{\text{п}}$ существенно разнятся.

Так, при $\text{random}_{\min} = 10^{-4}$ для распределения Парето при $m = 10$ и $\alpha = 1,1$ имеем $y_{\max}^{\text{п}} = \frac{1,618}{1,1\sqrt{10^{-4}}} = 17\,000,3$; при $\alpha = 2$ $y_{\max}^{\text{п}} = 505$.

Для экспоненциального распределения $y_{\max}^{\text{э}} = -m \ln 10^{-4} = 92,2$.

При увеличении диапазона дискретности эта разница в возможных выборочных максимальных значениях еще возрастает.

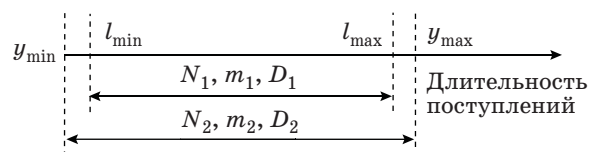
Получается, что такое прямое сопоставление значений очередей в СМО $M|M|1$ и $M|Pa|1$ некорректно.

Максимальные значения сообщений одного и того же реального трафика априори отображаются разными значениями при распределении Парето и при экспоненциальном распределении длительности поступлений: при одном и том же датчике случайных чисел заведомо $y_{\max}^{\text{п}} > y_{\max}^{\text{э}}$.

В табл. 4 представлены статистические характеристики Парето- ($\alpha = 1,1, k = 5$) и экспоненциально распределенных длительностей поступлений при ограничении, учитывающем возможный диапазон $[l_{\min}, l_{\max}]$ размеров поступающих сообщений. Такой подход имеет смысл из-за существования верхнего предела на размеры передаваемых кадров.

Приведенные в табл. 4 данные получены в результате численного эксперимента (рис. 3) с имитационными моделями СМО $M/Pa/1$ и $M/M/1$.

Численный эксперимент проводился при следующих объемах выборочных данных: $N_1 = 10^6$ — для диапазона $[l_{\min}, l_{\max}]$, N_2 — для диапазона $[y_{\min}, y_{\max}]$ (графа 2 при распределении Парето, графа 6 при экспоненциальном распределении). В графах 3 и 7 представлены отношения оценочных дисперсий к квадратам средних значений размеров поступлений при распределении Парето и экспоненциальном распределении соответственно без ограничения значений random, в графах 4 и 8 — те же отношения при ограничении минимального значения random (10^{-4}).



■ **Рис. 3.** Схема эксперимента

■ **Таблица 4**

№ п/п	$M Pa 1$				$M M 1$			
	$[l_{\min}, l_{\max}]$	N_2	$\frac{D_2}{m_2^2}$	$\frac{D_1}{m_1^2}$	$[l_{\min}, l_{\max}]$	N_2	$\frac{D_2}{m_2^2}$	$\frac{D_1}{m_1^2}$
	$y_{\max}^{\text{п}}$				$y_{\max}^{\text{э}}$			
	1	2	3	4	5	6	7	8
1	[5, 50] 1 855 032	1 086 696	3317,5	0,53	[5, 50] 603	1 703 447	1,001	0,29
2	[5, 100] 2 803 142	1 038 484	5680,1	1,01	[5, 100] 567	1 262 604	0,995	0,45
3	[5, 1000] 2 438 309	1 002 970	4668,8	5,97	[5, 1000] 527	1 126 757	1,001	0,80
4	[5, $2,5 \times 10^4$] 14 438 902	1 000 086	7039,5	63,76	[5, $2,5 \times 10^4$] 809	1 095 587	1,003	0,84
5	[5, 5×10^4] 900 547,5	1 000 044	1864,9	105,95	[5, 5×10^4] 572	1 130 892	1,001	0,79

Среднее значение m_2 размера поступлений при экспоненциальном распределении длительностей поступлений взято равным статистической оценке m_2 распределения Парето длительностей поступлений с тем, чтобы сохранить одинаковую нагрузку при обслуживании поступающих требований.

Соотношения средних значений очередей $L_{c,п}$ ($M|Pa|1$) и $L_{c,э}$ ($M|M|1$), вычисленные с использованием формулы Полячека — Хинчина и статистик (графы 3 и 7 соответственно) без учета калибровки моделей потоков к диапазону $[l_{\min}, l_{\max}]$, дали следующие результаты:

- вариант 1 — 1659,0;
- вариант 2 — 2913,4;
- вариант 3 — 2334,9;
- вариант 4 — 3467,9;
- вариант 5 — 1426,6.

В табл. 5 представлены соотношения средних и максимальных значений очередей тех же СМО, соответствующие выборочным данным при $y_{\max}^п \geq l_{\max}$ и $y_{\max}^э \geq l_{\max}$ (табл. 4, № п/п 1 и 2) и $y_{\max}^п < l_{\max}$ и $y_{\max}^э < l_{\max}$ (табл. 4, № п/п 3, 4 и 5) диапазона $[l_{\min}, l_{\max}]$.

При перекрытии диапазонов как при экспоненциальном распределении, так и при Парето отличие характеристик очередей процентное (графы 1 и 2 табл. 5), при перекрытии диапазонов при Парето-распределении и неперекрытии при экспоненциальном (графы 3, 4 и 5 табл. 5) отличие возрастает по мере увеличения различия в выборочных максимальных значениях.

Анализ представленных в табл. 5 результатов позволяет сделать следующие выводы, связанные с особенностью моделирования и сопоставления очередей $M|M|1$ и $M|Pa|1$.

Различие в характеристиках очереди между экспоненциальным обслуживанием и обслуживанием по Парето возрастает с увеличением нагрузки при $\alpha = 1,1$. При $\alpha \geq 1,5$ различие от нагрузки значительно менее заметно и существенно снижается при увеличении значения α (см. табл. 3).

Еще в большей степени выравнивание названных характеристик очередей заметно при приведении выборочных значений экспоненциально и Парето-распределенных случайных величин к возможному диапазону размеров сообщений реального трафика.

■ Таблица 5

Отношение очередей	№ варианта				
	1	2	3	4	5
$\frac{L_{c,п}}{L_{c,э}}$	1,25	1,34	3,83	35,2	56,0
$\frac{L_{м,п}}{L_{м,э}}$	1,04	1,43	3,48	38,4	46,1

Однако калибровка распределения Парето к диапазону длительностей протокольных единиц трафика есть не что иное, как «обрубание хвостов» распределения Парето. Распределение Парето с «хвостами», обрубленными по максимумам выборки ограниченного объема, отличается от неусеченного распределения Парето принципиально (особенно при бесконечных дисперсиях).

В физическом смысле такое усечение распределения Парето оправдано. Но насколько такое усечение корректно, и является ли «усеченное» Парето распределением с тяжелыми «хвостами»?

Итак, общих аналитических результатов исследования очередей или влияния самоподобности трафика на качество обслуживания пользователей телекоммуникационной сети в настоящее время не существует [9]. Привычные методы имитационного моделирования классических СМО при моделировании фрактальных СМО могут становиться несостоятельными. Причиной этого является проблематичность корректной реализации распределения с тяжелыми «хвостами», в частности, это продемонстрировано на распределении Парето [15].

Заключение

Открытие фрактальных свойств у трафика инфокоммуникационных сетей привело к радикальной корректировке математических моделей трафика и методов его обслуживания. При описании и анализе фрактального трафика стали широко применяться такие математические понятия, как самоподобный случайный процесс, долговременная зависимость, распределения с тяжелыми «хвостами». Фрактальные свойства сетевого трафика диктуют необходимость коренным образом пересмотреть модели расчета буферных накопителей.

Общих результатов аналитического анализа фрактальных СМО не существует. Известны лишь отдельные результаты для частных случаев. Исследование свойств распределений с тяжелыми «хвостами» посредством имитационного моделирования показывает, что такие распределения реализуются в общем случае со значительными искажениями. Это может приводить к существенным ошибкам в результатах моделирования фрактальных систем с очередями. Причиной искажения распределений с тяжелыми «хвостами» является дискретность используемых генераторов стандартных случайных чисел.

При сопоставлении результатов численных экспериментов, полученных на имитационной модели фрактальной и классической СМО при использовании одного и того же датчика случайных чисел, сопоставлять значения очередей, видимо, нельзя. И фрактальная, и экспоненциальная случайные величины распределены до бесконечных

значений. Но один и тот же датчик случайных чисел выдает максимальное значение для Парето-распределенной величины во много раз большее, чем для экспоненциальной (по значению, не по частоте). Это максимальное значение приходится либо нормировать по отношению к возможному диапазону значений передаваемых сообщений, либо калибровать. Однако это утверждение требует дополнительного исследования.

Перспективным представляется исследование многоканальных фрактальных СМО как моделей буферных накопителей сетевых узлов совместно с протоколом транспортного уровня SCTP (Stream Control Transmission Protocol — протокол пере-

дачи с управлением потоком). Этот протокол реализует многопоточность. Термин «многопоточность» (multi-streaming) обозначает способность протокола SCTP параллельно передавать данные по нескольким независимым потокам сообщений. Представляется, что многопоточность очень сильно коррелирует с параллельной обработкой фрактального потока. Другими словами, поступающий на узел фрактальный поток можно параллельно обрабатывать и в виде отдельных сообщений (отдельных потоков сообщений) передавать по нескольким исходящим каналам. Такое решение представляется эффективным, но также требует существенного исследования.

Литература

1. Шелухин О. И., Осин А. В., Смольский С. М. Самоподобие и фракталы. Телекоммуникационные приложения. — М.: Физматлит, 2008. — 368 с.
2. Костромицкий А. И., Волотка В. С. Подходы к моделированию самоподобного трафика // Восточно-Европейский журнал передовых технологий. 2010. № 46. С. 46–49.
3. Петров В. В. То, что вы хотели знать о самоподобных процессах, но стеснялись спросить. — М.: Радиотехника, 2003. — 112 с.
4. William Stallings. Data and Computer Communications. — Prentice Hall, 1997. — 798 p.
5. Tanenbaum A. Computer Networks. 4th Edition. — Prentice Hall, 2002. — 912 p.
6. Кутузов О. И., Марфин С. А. Сравнительный анализ очереди при самоподобной нагрузке // Изв. СПбГЭТУ «ЛЭТИ». 2014. № 3. С. 20–23.
7. Кутузов О. И., Сергеев В. Г., Татарникова Т. М. Коммутаторы в корпоративных сетях. Моделирование и расчет. — СПб.: Судостроение, 2003. — 170 с.
8. Кутузов О. И., Татарникова Т. М. Моделирование систем и сетей телекоммуникаций. — СПб.: РГТУ, 2012. — 134 с.
9. Шелухин О. И., Тенякшев А. М., Осин А. В. Фрактальные процессы в телекоммуникациях. — М.: Радиотехника, 2003. — 480 с.
10. Кутузов О. И., Татарникова Т. М. Математические схемы и алгоритмы моделирования инфокоммуникационных систем. — СПб.: ГУАП, 2013. — 148 с.
11. Кутузов О. И., Татарникова Т. М. Инфокоммуникационные сети. Моделирование и оценка вероятностно-временных характеристик. — СПб.: ГУАП, 2015. — 382 с.
12. Кутузов О. И., Хаддад М. Аналитико-статистический метод расчета малых вероятностей потерь в буфере конечной емкости // Телекоммуникационные технологии. 1994. Вып. 1. С. 36–48.
13. Задорожный В. Н., Кутузов О. И. Методы имитационного моделирования фрактальных систем с очередями // Материалы 6-й Всерос. НПК. Омск: ОмГТУ, 2015. С. 9–26.
14. Zwart A. P. Queueing Systems with Heavy Tails. — Eindhoven University of Technology, 2001. — 227 p.
15. Тонг Минь Дык, Кутузов О. И. Оценка влияния самоподобной нагрузки на характеристики очереди // Изв. СПбГЭТУ «ЛЭТИ». Сер. Информатика, управление и компьютерные технологии. 2006. Вып. 3. С. 35–38.

UDC 004.7

doi:10.15217/issn1684-8853.2016.2.48

Evaluation and Comparison of Queues in Classical and Fractal Queueing Systems

Kutuzov O. I.^a, Dr. Sc., Tech., Professor, kutuzov-oleg@mail.ru

Tatarnikova T. M.^b, Dr. Sc., Tech., Associate Professor, tm-tatarn@yandex.ru

^aSaint Petersburg Electrotechnical University «LETI», 5, Professora Popova St., 197376, Saint-Petersburg, Russian Federation

^bSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

Introduction: Buffer storages of network nodes are the most important traffic management resource. Research over the past decade and a half showed that network traffic is inherently self-similar and therefore the traditional queue analysis based on the Poisson flow assumption cannot accurately predict the performance of a node and the network in general. The lack of general analytical results

in studying queues with self-similar traffic poses the necessity to find a way of its correct description and modeling. **Purpose:** The goal is to develop a revenue model in self-similar traffic which would allow you to evaluate the queueing characteristics of classical and fractal queueing systems, and to compare them in the first approximation. **Results:** A model of revenues in self-similar traffic is proposed, characterized by the presence of long-term dependencies in the traffic routes and the presence of a pulsating structure on many time scales. This can adequately reflect the characteristics of the network traffic. Models are proposed for calculating the buffer storages of network nodes which are more realistic for self-similar traffic. **Practical relevance:** The revenue model for self-similar traffic and the model of calculating buffer storages can be applied in network traffic management and resource planning.

Keywords — Self-Similar Traffic, Bursty Traffic Pattern, ON/OFF-model, Long-Term Relationship, Distribution with Heavy “Tails”, Pareto Distribution, System Queues, Buffer Memory, Probability of Losses, Pollaczek–Khinchine formula, Limited Queue, Unlimited Queue, Network Traffic Management.

References

1. Shelukhin O. I., Osin A. V., Smol'skij S. M. *Samopodobie i fraktaly. Telekomunikacionnye prilozhenija*. [Self-similarity and Fractals. Telecommunication Applications]. Moscow, Fizmatlit Publ., 2008. 368 p. (In Russian).
2. Kostromickij A. I., Volotka V. S. Approaches to Modeling Self-similar Traffic. *Vostochno-Evropeiskii zhurnal pere-dovykh tekhnologii* [Eastern-European Journal of Enterprise Technologies], 2010, no. 46, pp. 46–49 (In Russian).
3. Petrov V. V. *Chto vy hoteli znat' o samopodobnykh processah, no stesnjalis' sprositi'* [What You Wanted to Know about Self-similar Processes, but Hesitated to Ask]. Moscow, Radiotekhnika Publ., 2003. 112 p. (In Russian).
4. William Stallings. *Data and Computer Communications*. Prentice Hall, 1997. 798 p.
5. Tanenbaum A. *Computer Networks*. 4th Edition. Prentice Hall, 2002. 912 p.
6. Kutuzov O. I., Marfin S. A. Comparative Analysis of the Queue when Self-Similar Load. *Izvestiia GETU "LETI"*, 2014, no. 3, pp. 20–23 (In Russian).
7. Kutuzov O. I., Sergeev V. G., Tatarnikova T. M. *Kommutator v korporativnykh setyakh. Modelirovanie i raschet* [Switches in Corporate Networks. Modeling and Calculation]. Saint-Petersburg, Sudostroenie Publ., 2003. 170 p. (In Russian).
8. Kutuzov O. I., Tatarnikova T. M. *Modelirovanie sistem i setey telekommunikatsiy* [Simulation of Systems and Telecommunications Networks]. Saint-Petersburg, RGGMU Publ., 2012. 134 p. (In Russian).
9. Shelukhin O. I., Osin A. V., Smol'skij S. M. *Fraktal'nye processy v telekommunikacijah* [Fractal Processes in Telecommunications]. Moscow, Radiotekhnika Publ., 2003. 408 p. (In Russian).
10. Kutuzov O. I., Tatarnikova T. M. *Matematicheskie shemy i algoritmy modelirovaniya infokommunikacionnykh sistem* [Mathematical Schemes and Communication Systems Simulation Algorithms]. Saint-Petersburg, GUAP Publ., 2013. 148 p. (In Russian).
11. Kutuzov O. I., Tatarnikova T. M. *Infokommunikacionnye seti. Modelirovanie i ocenka veroyatnostno-vremennykh harakteristik* [Infocommunication Network. Simulation and Evaluation of Probability-time Characteristics]. Saint-Petersburg, GUAP Publ., 2013. 148 p. (In Russian).
12. Kutuzov O. I., Haddad M. Analytical and Statistical Method for Calculating the Probability of Small Losses in the Buffer of Limited Capacity. *Telekommunikacionnye tekhnologii*, 1994, no. 1, pp. 36–48 (In Russian).
13. Zadorozhnyj V. N., Kutuzov O. I. Simulation Methods of Fractal Systems with Queues. *Materialy 6-j Vserossijskoj NPK* [Materials of the 6th All-Russian SPC]. Omsk, 2015, OmGTU Publ., pp. 9–26 (In Russian).
14. Zwart A. P. *Queueing Systems with Heavy Tails*. Eindhoven University of Technology, 2001. 227 p.
15. Tong Min Dyk, Kutuzov O. I. Assessing the Impact of Self-Similar Load on the Line Characteristics. *Izvestiia GETU "LETI". Ser. Informatika, upravlenie i komp'iuternye tekhnologii*, 2006, no. 3, pp. 35–38 (In Russian).

ПАМЯТКА ДЛЯ АВТОРОВ

Поступающие в редакцию статьи проходят обязательное рецензирование.

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail (ius.spb@gmail.com).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию. Рукописи не возвращаются.

Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.

ТЕОРЕТИКО-ИНФОРМАЦИОННЫЙ АНАЛИЗ МНОГОУРОВНЕВОЙ FLASH-ПАМЯТИ

Часть 1: Модель канала и границы случайного кодирования

А. Н. Трофимов^а, канд. техн. наук, доцент

Ф. А. Таубин^а, доктор техн. наук, профессор

^аСанкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ

Введение: flash-память представляет собой один из наиболее быстро растущих сегментов глобальной полупроводниковой индустрии. Благодаря неуклонно возрастающей плотности записи, высокой скорости записи/считывания, низкому энергопотреблению и продолжительному сроку службы, flash-память используется для хранения данных в весьма обширной сфере приложений. Повышение плотности записи, достигаемое за счет уменьшающегося физического размера ячейки наряду с возрастающим количеством используемых состояний, приводит к снижению надежности хранения данных, что требует использования помехоустойчивого кодирования. **Цель:** исследование теоретически достижимых значений основных параметров помехоустойчивого кодирования для модели многоуровневой NAND flash-памяти в зависимости от условий записи и хранения данных. **Результаты:** получены теоретически допустимые пределы скорости кодирования, для которых возможно построение надежных систем хранения данных для одной из возможных моделей flash-памяти, которая описывает страницу чипа памяти как систему с независимыми многоуровневыми ячейками. Для этой модели находятся границы случайного кодирования в форме, допускающей получение численных результатов. С использованием полученных границ устанавливаются обменные соотношения между предельно возможной скоростью кодирования и такими ключевыми параметрами, как число циклов перезаписи и время хранения данных. Рассмотрена возможность гауссовой аппроксимации канала записи/считывания. Эта аппроксимация оказывается в ряде случаев достаточно точной, что позволило получить точные оценки вычислительной скорости R_0^* в явном виде как функций от параметров гауссовых аппроксимаций. Показано, что во многих случаях величины пропускной способности и вычислительной скорости оказываются близкими — разница между ними составляет величину от долей процента до нескольких процентов. **Практическая значимость:** благодаря полученным результатам можно достаточно точно оценить степень снижения предельно достижимой плотности записи в многоуровневой NAND flash-памяти, связанную с ростом числа циклов перезаписи и увеличением длительности хранения данных. Это дает возможность, в частности, указать такие граничные значения числа циклов перезаписи и длительности хранения данных, при которых обеспечивается требуемая предельно достижимая плотность записи.

Ключевые слова — многоуровневая NAND flash-память, канал записи, теорема кодирования, пропускная способность, граница случайного кодирования.

Введение

Исследования и разработки, направленные на совершенствование flash-памяти — доминирующей технологии организации энергонезависимых устройств хранения данных, — образуют в настоящее время один из наиболее быстро растущих сегментов глобальной полупроводниковой индустрии. Благодаря неуклонно возрастающей плотности записи, высокой скорости записи/считывания, низкому энергопотреблению и продолжительному сроку службы, flash-память используется для хранения данных в весьма обширной сфере приложений — от потребительской электроники до корпоративных информационных систем [1, 2].

Основным элементом хранения данных во flash-памяти является ячейка, представляющая собой транзистор с плавающим затвором. Ячейка может находиться в одном из q ранжированных по уровню состояний: $0, 1, \dots, q-1$, — определяемых величиной заряда плавающего затвора и представляющих фактически некоторые поро-

вые уровни напряжения, которые могут быть измерены. Операция измерения порогового уровня напряжения в ячейке называется чтением (cell reading). Перевод ячейки из нулевого состояния в заданное ненулевое состояние (или из некоторого ненулевого состояния в более высокое состояние) осуществляется вводом в плавающий затвор (туннелированием) соответствующего заряда; эта операция называется записью, или программированием (cell writing or programming). Ячейка переводится в нулевое состояние путем удаления заряда из плавающего затвора; эта операция называется стиранием (cell erasing). Отметим, что для перевода ячейки из некоторого ненулевого состояния в более низкое ненулевое состояние необходимо вначале выполнить операцию стирания, после чего ввести в плавающий затвор заряд соответствующей величины. Flash-память с $q > 2$ называется многоуровневой (в отличие от одноуровневой памяти, где возможен лишь один ненулевой уровень). В силу более высокой плотности записи многоуровневая flash-память оказывается, очевидно, более предпочтительной

по сравнению с одноуровневой. В настоящее время стандартом де-факто является flash-память с $q = 4$. Вместе с тем все большее распространение получает восьмиуровневая flash-память [3, 4]; при этом в литературе сообщается о дальнейшем прогрессе технологии производства flash-памяти, позволяющем использовать 16 уровней [5].

Оборотной стороной повышения плотности записи, достигаемого за счет неуклонно уменьшающегося физического размера ячейки в совокупности с возрастающим количеством используемых состояний, является снижение надежности хранения данных. К основным факторам, определяющим снижение надежности, обычно относят следующие эффекты, характерные для flash-памяти. Во-первых, это случайные изменения величины заряда, вводимого в плавающий затвор транзистора в процессе записи (program disturbance), что проявляется в отклонении реального порогового уровня напряжения от целевого уровня. Во-вторых, повторяющиеся в процессе эксплуатации циклы записи/стирания (program/erase, P/E cycling), каждый из которых приводит к небольшому физическому разрушению ячейки; эти разрушения проявляются, в частности, в появлении ловушек вводимого/выводимого заряда (charge traps). Следующий фактор — паразитная емкостная связь между соседними транзисторами, вызывающая взаимную интерференцию ячеек (cell-to-cell interference). Наконец, четвертым фактором является утечка заряда плавающего затвора с течением времени (retention problem), что снижает исходный пороговый уровень. Совокупность указанных факторов приводит к случайному отклонению реального порогового уровня напряжения от целевого уровня (target value) напряжения в ячейке.

В результате в процессе считывания содержимого flash-памяти возникают ошибки. Экспериментальные исследования [6, 7] показывают, что в многоуровневой памяти доминирующим является распределение ошибок, близкое к равномерному на множестве ячеек, при этом не наблюдается заметной тенденции к пакетированию. Последнее обстоятельство может служить аргументом для формального описания рассматриваемого канала записи как *канала без памяти*. Учитывая, что вероятность ошибки (raw bit error rate) в многоуровневой flash-памяти оказывается неприемлемо высокой — порядка 10^{-4} и более, тогда как требуемая вероятность ошибки лежит в диапазоне $10^{-12} \dots 10^{-16}$, введение помехоустойчивого кодирования становится неизбежным. Одна из существенных особенностей использования помехоустойчивого кодирования во flash-памяти состоит в необходимости согласо-

вывать рассматриваемый вариант кодирования с архитектурой чипа flash-памяти.

Существуют два основных типа flash-памяти — NOR flash-память и NAND flash-память. В NOR flash-памяти возможен доступ в произвольном порядке к любой ячейке для выполнения операций записи и (или) чтения, тогда как в NAND flash-памяти эти операции могут выполняться лишь одновременно со всеми ячейками, входящими в некоторую группу, называемую страницей. Другими словами, для NAND flash-памяти характерны определенные ограничения при организации процедуры записи, однако это с избытком компенсируется более высокой (по сравнению с NOR flash-памятью) достижимой плотностью записи. В дальнейшем рассмотрении ограничимся исследованием модели NAND flash-памяти.

Конструктивно чип NAND flash-памяти состоит, как правило, из нескольких десятков расположенных друг над другом слоев двумерных массивов ячеек, называемых блоками. Блок есть своего рода неделимая единица при выполнении операции стирания (block granularity), т. е. в нулевое состояние переводятся одновременно все ячейки блока. Каждый блок делится на страницы, которые являются минимальными (по размеру) адресуемыми элементами при выполнении операций записи и (или) чтения (page granularity). В большинстве случаев типичная страница включает от 2 К до 64 К ячеек и делится на две области: в первой области хранятся собственно данные, а вторая, резервная область (spare area), предназначена, в частности, для проверочных символов помехоустойчивого кода. Как правило, доля резервной области составляет 3–7 %. При сравнительном анализе различных вариантов помехоустойчивого кодирования для flash-памяти важным с прикладной точки зрения представляется использование теоретико-информационного инструментария. Наличие такого инструментария позволяет, например, оценить, насколько близка реализуемая скорость кодирования (эта величина, очевидно, напрямую связана с реальной плотностью записи) в том или ином случае к предельно возможным значениям. Один из подходов к оцениванию теоретико-информационных границ для предельно возможной скорости кодирования представлен в работах [8, 9]. В данной работе на основе (одной из принятых в литературе) формализованной модели канала записи-считывания находятся границы случайного кодирования в форме, допускающей получение численных результатов. С использованием построенных границ устанавливаются обменные соотношения между предельно возможной скоростью кодирования и такими ключевыми параметрами, как число циклов перезаписи и время хранения данных.

Модель канала многоуровневой flash-памяти

Как отмечалось во введении, минимальными (по объему памяти) адресуемыми элементами при выполнении операций записи и (или) чтения являются страницы чипа. Это означает, что все ячейки страницы находятся в одинаковых условиях (с точки зрения влияния числа циклов записи/стирания и времени хранения), поэтому физическая среда, соответствующая странице, может рассматриваться как стационарный канал. С учетом отсутствия заметной тенденции к пакетированию ошибок [6] будем полагать, что рассматриваемый канал представляет собой канал без памяти. В результате модель канала полностью определяется моделью одиночной ячейки flash-памяти. Опишем кратко упрощенную математическую модель одиночной ячейки flash-памяти в соответствии с имеющимися публикациями [8, 9].

Ячейка многоуровневой flash-памяти характеризуется q уровнями входного сигнала (напряжения).

Пример. Распространенным примером, используемым далее в качестве основного, служит значение $q = 4$ [8, 9, 10]. ■

Каждый из уровней выходного значения соответствует записи одной из m -разрядных двоичных комбинаций, где $m = \log_2 q$. Уровни представляют собой выходные величины, изменяющиеся от ячейки к ячейке, и поэтому могут рассматриваться как случайные величины, определенные на множестве ячеек, образующих блок flash-памяти.

Как указывалось во введении, имеются следующие факторы, влияющие на распределение значений q выходных уровней одиночных ячеек: а) начальное распределение пороговых значений; б) влияние циклов записи/стирания; в) взаимная интерференция ячеек; г) влияние времени хранения.

Оставляя в стороне описание физических процессов, влияющих на распределение выходных уровней одиночной ячейки flash-памяти, дадим ее описание на уровне сравнительно простой математической модели.

Пусть $x_0, x_1, \dots, x_{q-1}$ — некоторым образом нормализованные значения *номинальных*, или *начальных*, уровней выходного сигнала ячейки.

Пример (продолжение). В рассматриваемом примере значения номинальных, или начальных, уровней $x_0, x_1, \dots, x_3$ равны $x_0 = 1,4, x_1 = 2,6, x_2 = 3,2, x_3 = 3,93$ [8, 9]. ■

Значения выходных уровней ячейки представляют собой случайные величины, распределения которых параметризованы значениями $x_0, x_1, \dots, x_{q-1}$, т. е. они описываются условными

функциями плотности вероятности $p_{y|x}(y|x)$, где $-\infty < y < \infty, x = x_0, x_1, \dots, x_{q-1}$. Условные функции плотности вероятности $p_{y|x}(y|x)$ задаются следующим образом:

$$p_{y|x}(y|x) = \begin{cases} p_0(y|x), & \text{если } x = x_0; \\ p_0(y|x) * p_r(y) * p_c(y) * p_d(y), & \text{если } x = x_1, x_2, \dots, x_{q-1}. \end{cases} \quad (1)$$

Здесь $p_0(\cdot)$ — функции плотности вероятности, задающие распределения *номинальных* выходных значений; $p_r(\cdot), p_c(\cdot), p_d(\cdot)$ — функции плотности вероятности, определяющие соответственно влияние записи/стирания, взаимную интерференцию ячеек и влияние времени хранения; знак $*$ обозначает операцию свертки. Как следует из выражения (1), имеется один уровень выходного сигнала x_0 , который не подвержен влиянию иных мешающих факторов. Остальные выходные уровни $x_1, \dots, x_{q-1}$ подвергаются дополнительным искажениям, которые описываются как аддитивные независимые случайные слабые.

Функции плотности вероятности $p_0(\cdot)$, задающие распределения начальных уровней, определены как гауссова плотность для $x = x_0$:

$$p_0(y|x_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x_0)^2}{2\sigma^2}}, \quad (2)$$

и как равномерные плотности для $x = x_1, \dots, x_{q-1}$, т. е.

$$p_0(y|x_i) = \begin{cases} \frac{1}{\Delta_i}, & \text{если } |y - x_i| \leq \frac{\Delta_i}{2}; \\ 0 & \text{в противном случае,} \end{cases} \quad (3)$$

где $i = 1, 2, \dots, q - 1$.

Пример (продолжение). В рассматриваемом примере [7, 8] значения параметров распределений начальных уровней задаются как $\sigma = 0,35, \Delta_1 = \Delta_2 = \Delta_3 = 0,2$. ■

Рассмотрим теперь влияние остальных мешающих факторов. Функция плотности вероятности $p_r(\cdot)$, определяющая влияние циклов записи/стирания, задается как двусторонняя симметричная экспоненциальная плотность:

$$p_r(y) = \frac{1}{2\lambda_r} e^{-\frac{|y|}{\lambda_r}}, \quad (4)$$

где параметр λ_r связан с числом циклов записи/стирания N зависимостью $\lambda_r = K_\lambda \sqrt{N}$, K_λ — постоянный коэффициент.

Пример (продолжение). В основном примере $K_\lambda = 0,00025$. ■

Функция плотности вероятности $p_c(\cdot)$, определяющая взаимную интерференцию ячеек, в рас-

смаатриваемой модели задана как усеченная гауссова плотность, т. е. она имеет вид

$$p_c(y) = \begin{cases} \frac{c}{\sqrt{2\pi}\sigma_c} e^{-\frac{(y-\mu_c)^2}{2\sigma_c^2}}, & \text{если } |y-\mu_c| \leq w_c; \\ 0 & \text{в противном случае,} \end{cases} \quad (5)$$

где c — нормирующий коэффициент, обеспечивающий выполнения равенства $\int_{-\infty}^{\infty} p_c(y) dy = 1$.

Пример (продолжение). В основном примере значения параметров распределения (5) указаны как $w_c = 0,1\mu_c$ и $\sigma_c = 0,4\mu_c$. Значение величины μ_c примем равным $\mu_c = 0,2$ исходя из условия согласования окончательных плотностей вероятности (1) с данными, приведенными в работах [8, 9]. ■

Функция плотности вероятности $p_d(\cdot)$, описывающая влияние времени хранения записанных данных (retention), задается как нормальная плотность вероятности

$$p_d(y) = \frac{1}{\sqrt{2\pi}\sigma_d} e^{-\frac{(y-\mu_d)^2}{2\sigma_d^2}}, \quad (6)$$

где величины μ_d и σ_d определяются как

$$\mu_d = -K_s(x - x_0)K_d N^{0,5} \ln(1 + T/T_0); \quad (7)$$

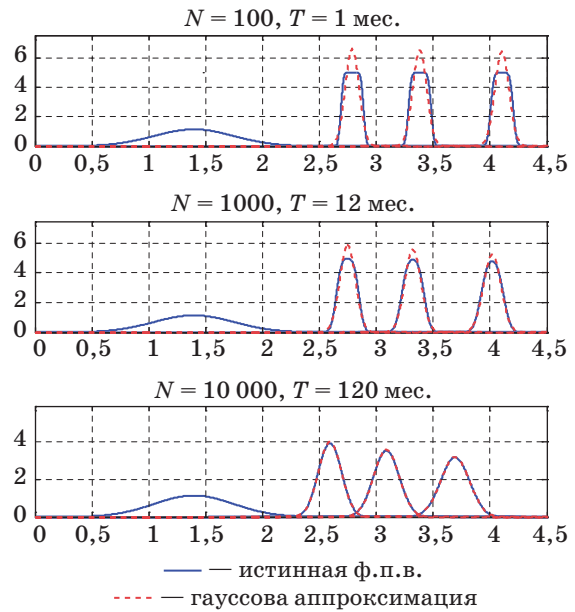
$$\sigma_d = K_s(x - x_0)K_m N^{0,6} \ln(1 + T/T_0). \quad (8)$$

В выражениях (7) и (8) N , как и ранее, обозначает число циклов перезаписи; T — время хранения; T_0 — интервал времени, соответствующий одному часу; K_s , K_d и K_m — постоянные коэффициенты. Заметим, что в работах [8, 9], по всей видимости, выражение для величины μ_d приведено с опечаткой — пропущен знак минус. При таком определении величина μ_d не убывает, а увеличивается с ростом числа циклов перезаписи. При этом выходит, что со временем характеристики канала улучшаются, а не ухудшаются, что не соответствует практике.

Пример (окончание). В основном примере постоянные коэффициенты K_s , K_d и K_m определены следующим образом: $K_s = 0,38$, $K_d = 4 \cdot 10^{-4}$ и $K_m = 4 \cdot 10^{-6}$. ■

Графики функций плотностей вероятности $p_{y|x}(y|x)$ для некоторых типичных примеров показаны на рис. 1. Эти графики получены путем численного нахождения сверток в формуле (1); они хорошо согласуются с аналогичными графиками из работ [8, 9], полученными в результате статистического моделирования (методом Монте-Карло).

Исходя из внешнего вида графиков (см. рис. 1) можно предположить, что приближение их гауссовыми плотностями с соответствующим образом подобранными параметрами будет приемлемым.



■ **Рис. 1.** Функции плотности вероятности $p_{y|x}(y|x)$, $x = x_0, \dots, x_3$, и их гауссовы приближения для различных значений параметров N и T

Эти приближения также показаны на рис. 1. Заметим, что модели канала flash-памяти, использующие гауссовы аппроксимации, также встречаются в публикациях [10].

Рассмотренная модель, определяемая функциями плотности вероятности $p_{y|x}(y|x)$ или их гауссовыми аппроксимациями, называемая также моделью с гауссовым шумом, стандартное отклонение которого зависит от входного значения (input-dependent additive Gaussian noise) [10], допускает обобщение на большее число входных уровней, в частности на шесть, восемь и двенадцать уровней [11]. Такое обобщение не представляется сложным и может быть при необходимости выполнено.

Пределы скорости кодирования для канала многоуровневой flash-памяти

В предыдущем разделе рассмотрена модель канала многоуровневой записи на flash-носитель. Модель канала определяется набором одномерных функций плотности вероятности выходного значения элемента flash-памяти $p_{y|x}(y|x)$, $x = x_0, x_1, \dots, x_{q-1}$, q — число уровней записи. Далее ограничимся представленным в предыдущем разделе основным примером с $q = 4$ и остальными параметрами, указанными там же. Канал, таким образом, задается четырьмя функциями плотности вероятности $p_{y|x}(y|x_0)$, $p_{y|x}(y|x_1)$, $p_{y|x}(y|x_2)$ и $p_{y|x}(y|x_3)$, а значения номинальных (начальных) уровней записи x_0, x_1, x_2, x_3 соответствуют парам записанных бит 00, 01, 10, 11. Взаимно-однозначное соответствие $\{00, 01, 10, 11\} \leftrightarrow \{x_0, x_1, x_2, x_3\}$

определяется *модуляционным отображением*. Распространенным примером такого отображения может служить код Грея, при котором ближайшим сигналам (уровням записи) соответствуют двоичные комбинации, отличающиеся в одном разряде.

Каждая функция плотности вероятности $p_{y|x}(y|x)$ описывает распределение выходного значения одного элемента flash-памяти, и величину $\int_{-\infty}^u p_{y|x}(y|x)dy$ можно интерпретировать как среднюю долю числа элементов flash-памяти, имеющих выходное значение, не превышающее u при условии записи в них значения x .

В этом разделе сначала приводятся общие соотношения, позволяющие вычислить теоретико-информационные пределы для произвольного канала без памяти, а потом этот известный подход применяется к модели канала, представленной в предыдущем разделе.

Граница случайного кодирования (граница Галлагера) [12] определяет предельные характеристики надежной передачи данных или надежного хранения данных при использовании помехоустойчивого кодирования. Эта граница может быть построена как для блочных, так и для решетчатых, в частности сверточных, кодов [13]. Далее будет рассматриваться случай блочного кодирования как наиболее подходящий для применения в канале записи данных.

Определим входной и выходной ансамбли канала X и Y . В рассматриваемом случае $X = \{x_0, x_1, \dots, x_{q-1}\}$, $p_x(\cdot)$ — распределение, заданное на множестве $\{x_0, x_1, \dots, x_{q-1}\}$; $Y = \{\mathbb{R}, p_y(\cdot)\}$, где $\mathbb{R}$ — множество вещественных чисел и $p_y(\cdot)$ — функция плотности вероятности, заданная на $\mathbb{R}$. Здесь рассматривается вариант канала с непрерывным выходом. Этот вариант позволяет обеспечить наибольшую достижимую надежность передачи (хранения) данных. Случай дискретного (квантованного) выхода канала также может быть рассмотрен. Квантование выхода канала неизбежно приводит к потерям, и оценка этих потерь может быть предметом отдельного исследования.

Очевидно, что

$$p_y(y) = \sum_x p_x(x) p_{y|x}(y|x).$$

Граница случайного кодирования формулируется для принятой модели как теорема кодирования для канала без памяти с непрерывным выходом.

Теорема кодирования [12]. Существует блочный код длины n со скоростью $R = \log_2 M/n$ [бит/символ канала]; M — число слов кода, для которого вероятность ошибки декодирования P_e кодового блока (слова) по максимуму правдоподобия удовлетворяет неравенству $P_e < 2^{-nE(R)}$,

где $E(R)$ — экспонента случайного кодирования, определенная как

$$E(R) = \max_{p_x} \max_{0 < \rho < 1} (E_0(\rho, p_x) - \rho R). \quad (9)$$

Здесь

$$E_0(\rho, p_x) = -\log_2 \int_{-\infty}^{\infty} \left(\sum_x p_x(x) p_{y|x}(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} dy \quad (10)$$

— функция Галлагера; $p_x(\cdot)$ — распределение, заданное на множестве входных символов канала (значений номинальных записанных уровней), определяющее ансамбль случайных кодов. ■

Максимальная допустимая скорость кода, при которой возможна надежная передача (хранение) данных, называется пропускной способностью канала и определяется как

$$C = \max_{p_x} I(X; Y), \quad (11)$$

где

$$\begin{aligned} I(X; Y) &= \sum_x \int_{-\infty}^{\infty} p_{xy}(x, y) \log_2 \frac{p_{xy}(x, y)}{p_x(x) p_y(y)} dy = \\ &= \sum_x \int_{-\infty}^{\infty} p_x(x) p_{y|x}(y|x) \log_2 \frac{p_{y|x}(y|x)}{\sum_{x'} p_x(x') p_{y|x'}(y|x')} dy \end{aligned}$$

— средняя взаимная информация ансамблей X и Y .

Известно [12], что $E(R) > 0$ для $R < C$. Кроме того, важное следствие из теоремы кодирования состоит в том, что пропускная способность канала (11) и функция Галлагера (10) связаны соотношением

$$C = \max_{p_x} \left. \frac{\partial E_0(p_x, \rho)}{\partial \rho} \right|_{\rho=0}. \quad (12)$$

Важной характеристикой канала является так называемая вычислительная скорость канала

$$\begin{aligned} R_0 &= \max_{p_x} E_0(p_x, 1) = \\ &= -\log_2 \min_{p_x} \int_{-\infty}^{\infty} \left(\sum_x p_x(x) \sqrt{p_{y|x}(y|x)} \right)^2 dy. \quad (13) \end{aligned}$$

Эта величина может рассматриваться как предельное значение *практически достижимой* скорости кода.

Известно [12], что для каналов без памяти справедливо неравенство $C/2 < R_0 < C$, и во многих практически важных случаях вычислительная скорость R_0 незначительно меньше пропускной способности C . Поэтому величина R_0 в этих случаях может служить адекватным приближением к пропускной способности, кроме того,

ее положительным свойством является и то, что она значительно проще вычисляется.

Наряду с максимально достижимыми величинами $E(R)$, R_0 и C рассматриваются также приближения к ним, которые получаются, если в формулах (9)–(13) исключить оптимизацию по распределению на множестве входных значений канала $p_x(\cdot)$ и положить его равным некоторому фиксированному распределению, например *равномерному*. В этом случае возникают величины $E^*(R) \leq E(R)$, $R_0^* \leq R_0$, $C^* \leq C$, вычисляемые по формулам

$$E^*(R) = \max_{0 < \rho < 1} (E_0^*(\rho) - \rho R); \quad (14)$$

$$E_0^*(\rho) = -\log_2 \int_{-\infty}^{\infty} \left(\frac{1}{q} \sum_x p_{y|x}(y|x) \frac{1}{1+\rho} \right)^{1+\rho} dy = \\ = (1+\rho) \log_2 q - \log_2 \int_{-\infty}^{\infty} \left(\sum_x p_{y|x}(y|x) \frac{1}{1+\rho} \right)^{1+\rho} dy; \quad (15)$$

$$R_0^* = E_0^*(1) = 2 \log_2 q - \int_{-\infty}^{\infty} \left(\sum_x \sqrt{p_{y|x}(y|x)} \right)^2 dy; \quad (16)$$

$$C^* = \left. \frac{\partial E_0^*(\rho)}{\partial \rho} \right|_{\rho=0}. \quad (17)$$

Выражения для R_0 и R_0^* (13) и (16) могут быть преобразованы к более удобному виду с использованием следующего обозначения. Пусть

$$D(x, x') = \int_{-\infty}^{\infty} \sqrt{p_{y|x}(y|x) p_{y|x'}(y|x')} dy \quad (18)$$

— коэффициент Бхаттачария [13] для распределений, определяемых плотностями $p_{y|x}(y|x)$ и $p_{y|x'}(y|x')$. Очевидно, что $D(x, x') = D(x', x)$ и $D(x, x) = 1$.

Тогда из (13) следует, что

$$R_0 = -\log_2 \min_{p_x} \sum_x \sum_{x'} p_x(x) p_x(x') D(x, x') = \\ = -\log_2 \min_{p_x} \left(\sum_x p_x^2(x) + 2 \sum_x \sum_{x' > x} p_x(x) p_x(x') D(x, x') \right), \quad (19)$$

и если положить в (19) $p_x(x) = 1/q$ для всех x , то

$$R_0^* = 2 \log_2 q - \log_2 \left(q + 2 \sum_x \sum_{x' > x} D(x, x') \right). \quad (20)$$

Оптимизация по распределению $p_x(\cdot)$ в правой части (19) может быть выполнена в явном виде [14] с использованием метода множителей

Лагранжа. Задача оптимизации в этом случае формулируется как

$$\sum_x \sum_{x'} p_x(x) p_x(x') D(x, x') \rightarrow \min, \\ \sum_x p_x(x) = 1, \quad p_x(x) \geq 0.$$

Решение для нее можно выразить следующим образом:

$$\sum_{x'} D(x, x') p_x(x') = \lambda, \quad (21)$$

где λ — множитель Лагранжа, значение которого находится из условия $\sum_x p_x(x) = 1$. При решении системы линейных уравнений (21) нужно также учесть требование $p_x(x) \geq 0$ для всех x . Из условия (21) получаем, что

$$\sum_x \sum_{x'} p_x(x) p_x(x') D(x, x') = \sum_x p_x(x) \lambda = \lambda,$$

и, следовательно, с учетом первого равенства в выражении (19) оптимизированное значение вычислительной скорости

$$R_0 = -\log_2 \lambda. \quad (22)$$

Оптимизация по распределению $p_x(\cdot)$ в выражении (11) для пропускной способности выполняется сложнее. Здесь можно применить алгоритм Блейхута — Аримото (см., например, [13]), где он сформулирован для канала с дискретным входом и дискретным выходом. Если представить канал с непрерывным выходом как дискретный канал с достаточно большим числом выходов, т. е. представить непрерывные по аргументу y плотности $p_{y|x}(y|x)$ как дискретные распределения с большим числом исходов, то указанный алгоритм может быть применен. Пусть $\tilde{p}_{y|x}(\tilde{y}|x)$ — дискретное распределение, соответствующее плотности $p_{y|x}(y|x)$, где $\tilde{y}$ — дискретное (квантованное) значение величины y . Строго говоря, при вычислениях по формуле (1) в результате и получаются распределения $\tilde{p}_{y|x}(\tilde{y}|x)$, а не функции плотности вероятности $p_{y|x}(y|x)$. Тогда можно записать, что приближением к C будет величина

$$\tilde{C} = \max_{p_x} I(X; \tilde{Y}), \quad (23)$$

где

$$I(X; \tilde{Y}) = \sum_x \sum_{\tilde{y}} p_x(x) \tilde{p}_{y|x}(\tilde{y}|x) \times \\ \times \log_2 \frac{\tilde{p}_{y|x}(\tilde{y}|x)}{\sum_{x'} p_x(x') \tilde{p}_{y|x'}(\tilde{y}|x')}$$

— средняя взаимная информация ансамблей X и $\tilde{Y} = \{\{\tilde{y}\}, \tilde{p}_{y|x}(\tilde{y}|x)\}$. Приближение $\tilde{C}$ к истинному значению C будет тем точнее, чем больше будет число уровней дискретного представления $\tilde{y}$

непрерывного выхода канала y . Алгоритм определения оптимального распределения $p_x(\cdot)$, максимизирующего правую часть (23), выглядит следующим образом [13]. Пусть

$$Q^{(k)}(x|\tilde{y}) = \frac{\tilde{p}_{y|x}(\tilde{y}|x)p_x^{(k)}(x)}{\sum_{x'} \tilde{p}_{y|x}(\tilde{y}|x')p_x^{(k)}(x')}; \quad (24a)$$

$$p_x^{(k+1)}(x) = \frac{\exp\left(\sum_{\tilde{y}} \tilde{p}_{y|x}(\tilde{y}|x) \ln Q^{(k)}(x|\tilde{y})\right)}{\sum_{x'} \exp\left(\sum_{\tilde{y}} \tilde{p}_{y|x}(\tilde{y}|x') \ln Q^{(k)}(x'|\tilde{y})\right)}; \quad (24б)$$

$$\begin{aligned} \tilde{C}^{(k+1)} = & -\sum_x p_x^{(k+1)}(x) \ln p_x^{(k+1)}(x) + \\ & + \sum_x \sum_{\tilde{y}} \tilde{p}_{y|x}(\tilde{y}|x) p_x^{(k+1)}(x) \ln Q^{(k)}(x|\tilde{y}). \end{aligned} \quad (24в)$$

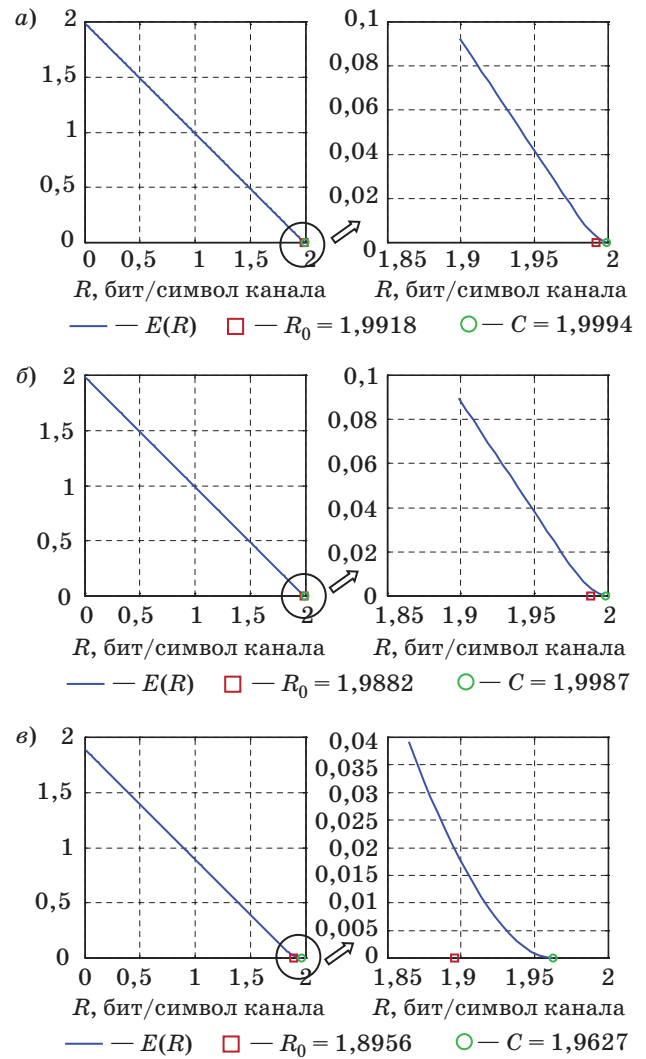
Тогда, начав с произвольного, например с равномерного, распределения $p_x^{(0)}(x)$ и выполняя вычисления (24) для $k = 0, 1, 2, \dots$, можно получить последовательность значений $\tilde{C}^{(1)}, \tilde{C}^{(2)}, \dots, \tilde{C}^{(k)}, \dots$, сходящуюся после очевидной замены единиц измерения с *нат/символ канала* на *бит/символ канала* к величине в правой части равенства (23).

Графики экспоненты случайного кодирования $E(R)$ и приближения к ней $E^*(R)$ имеют характерный вид линии, состоящей из двух частей, — криволинейной и прямолинейной, при этом криволинейная часть тем меньше, чем ближе значения R_0 и C , соответственно — R_0^* и C^* .

Примеры экспонент случайного кодирования $E(R)$, вычисленные согласно выражениям (9), (10), (12), (13) для некоторых вариантов канала записи/хранения данных, характеризуемых параметрами N и T , показаны на рис. 2, а–в.

Графики имеют вид прямых почти для всех значений скорости кода R , и значения R_0 почти совпадают с C . Кроме того, из результатов вычислений следует, что распределение $p_x(\cdot)$, максимизирующее величину R_0 , оказывается для рассмотренных примеров практически равномерным. То же самое относится и к распределениям $p_x(\cdot)$, максимизирующим значение средней взаимной информации в равенствах (11) и (23). Это значит, что значения R_0^* и C^* оказываются близкими к значениям R_0 и C соответственно. Поэтому в ряде случаев хорошей оценкой пропускной способности канала C может служить величина R_0 , а для нее хорошим оказывается приближение R_0^* , вычисляемое с использованием гауссовых аппроксимаций плотностей $p_{y|x}(y|x)$.

При использовании гауссовых приближений к функциям плотности вероятности $p_{y|x}(y|x)$ при



■ Рис. 2. Экспонента случайного кодирования $E(R)$: а — для $N = 100, T = 1$ мес.; б — для $N = 1000, T = 12$ мес.; в — для $N = 10\,000, T = 120$ мес. (слева — полный график, справа — часть графика для больших значений скорости кода R)

вычислении величин R_0 и R_0^* по формулам (19) и (20) полезным оказывается следующее утверждение.

Утверждение. Пусть

$$\begin{aligned} w_0(u) &= (2\pi\sigma_0^2)^{-1/2} \exp(-(u-m_0)^2/2\sigma_0^2) \text{ и} \\ w_1(u) &= (2\pi\sigma_1^2)^{-1/2} \exp(-(u-m_1)^2/2\sigma_1^2) \end{aligned}$$

— две гауссовы плотности вероятности с параметрами (m_0, σ_0^2) и (m_1, σ_1^2) соответственно. Тогда коэффициент Бхаттачария для этих плотностей равен

$$\int_{-\infty}^{\infty} \sqrt{w_0(u)w_1(u)} du = \sqrt{\frac{2\sigma_0\sigma_1}{\sigma_0^2 + \sigma_1^2}} \exp\left(-\frac{(m_0 - m_1)^2}{4(\sigma_0^2 + \sigma_1^2)}\right). \quad (25)$$

Равенство (25) доказывается непосредственным интегрированием.

По равенству (25) можно вычислять величины $D(x, x')$ в правых частях равенств (19) и (20) непосредственно с использованием значения параметров гауссовых аппроксимаций плотностей $p_{y|x}(\cdot)$.

Результаты вычислений значений R_0 , C , а также R_0^* и C^* , полученные с использованием истинных распределений с одной стороны и с использованием их гауссовых аппроксимаций с другой, оказываются очень близкими, разница имеется в третьем-четвертом знаке после точки. Это может служить аргументом в пользу использования гауссовых приближений к плотностям $p_{y|x}(\cdot)$ в расчетах параметров практических систем кодирования для каналов flash-памяти и предположения о равномерности распределения $p_x(\cdot)$.

Изменение значений пропускной способности C и вычислительной скорости R_0 при увеличении числа циклов перезаписи N при различных значениях времени хранения данных T показано на рис. 3, а и б. Там же показаны зависимости от N значений C и R_0 , полученные при использовании гауссовых приближений плотностей $p_{y|x}(\cdot)$.

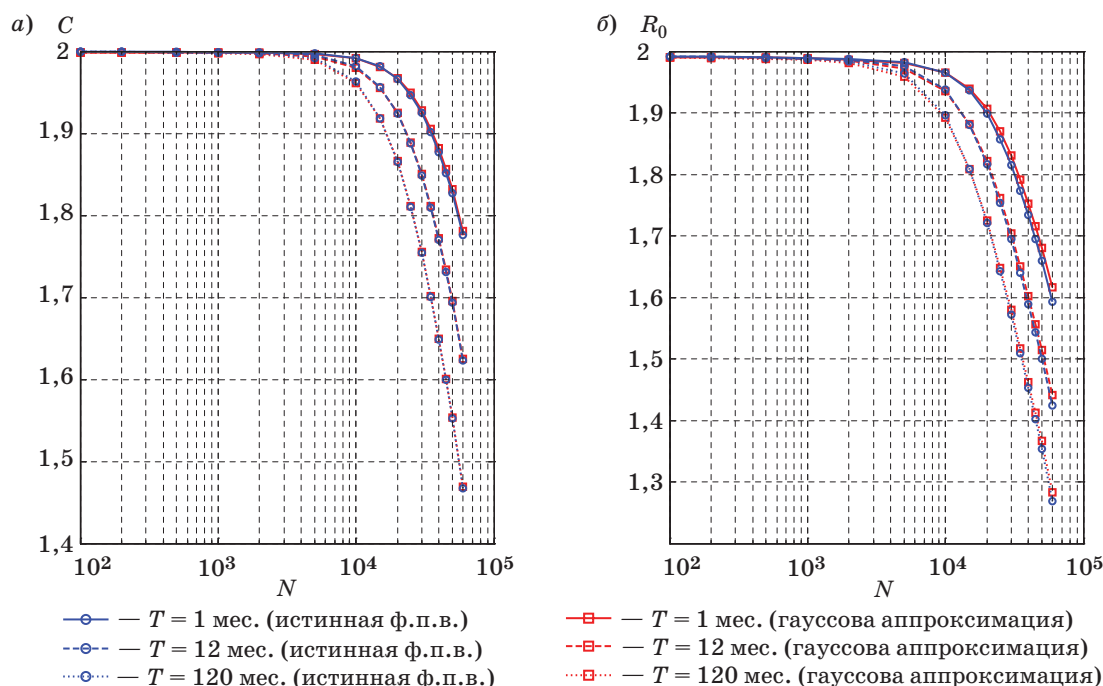
Как и следовало ожидать, значения C и R_0 уменьшаются с ростом N , поскольку при этом происходит ухудшение характеристик канала, которое выражается в «сближении» плотностей вероятности $p_{y|x}(\cdot)$ (см. рис. 1). Это ухудшение, однако, оказывается практически незаметным при увеличении числа циклов перезаписи N до значений порядка $(6...7) \cdot 10^3$. При дальнейшем увеличении N значения C и R_0 убывают гораздо

быстрее. Такой характер изменения наблюдается при всех рассмотренных значениях времени хранения данных T . Отметим, что величины C и R_0 , вычисленные для «истинных» плотностей $p_{y|x}(\cdot)$ и для их гауссовых приближений, оказываются очень близкими.

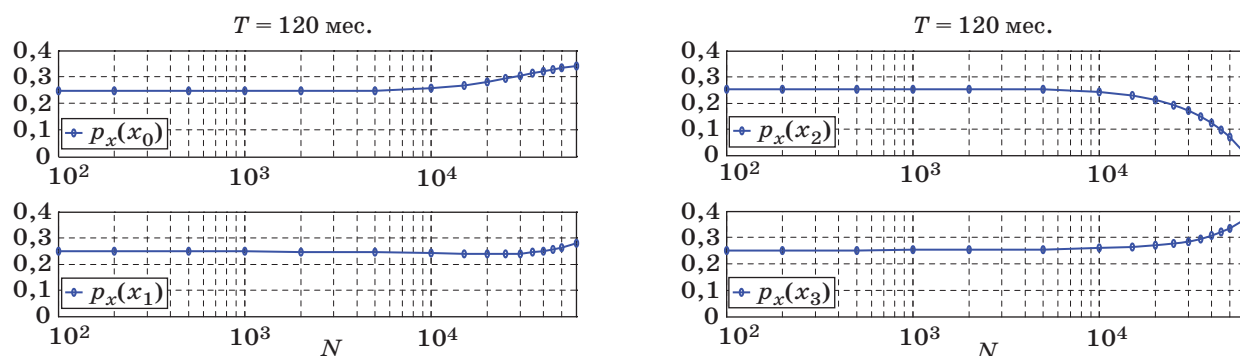
На рис. 4 представлены распределения $p_x(\cdot)$, задающие ансамбль случайных кодов, в зависимости от числа циклов перезаписи N . Эти распределения доставляют максимум величине R_0 , графики которой указаны на рис. 3, б.

Из рассмотрения рис. 4 следует, что оптимальное распределение $p_x(\cdot)$ остается практически равномерным ($p_x(\cdot) \approx 1/4$ для всех x) до значений N порядка 10^4 . При дальнейшем увеличении N распределения меняются и становятся заметно отличающимися от равномерного. При этом существенно уменьшается вероятность $p_x(x_2)$, что косвенно означает, что входной символ канала (уровень записи) x_2 становится очень ненадежным, и его следует исключить. Подчеркнем, что это явление наблюдается при очень большом числе циклов перезаписи N , когда характеристики канала заметно ухудшаются.

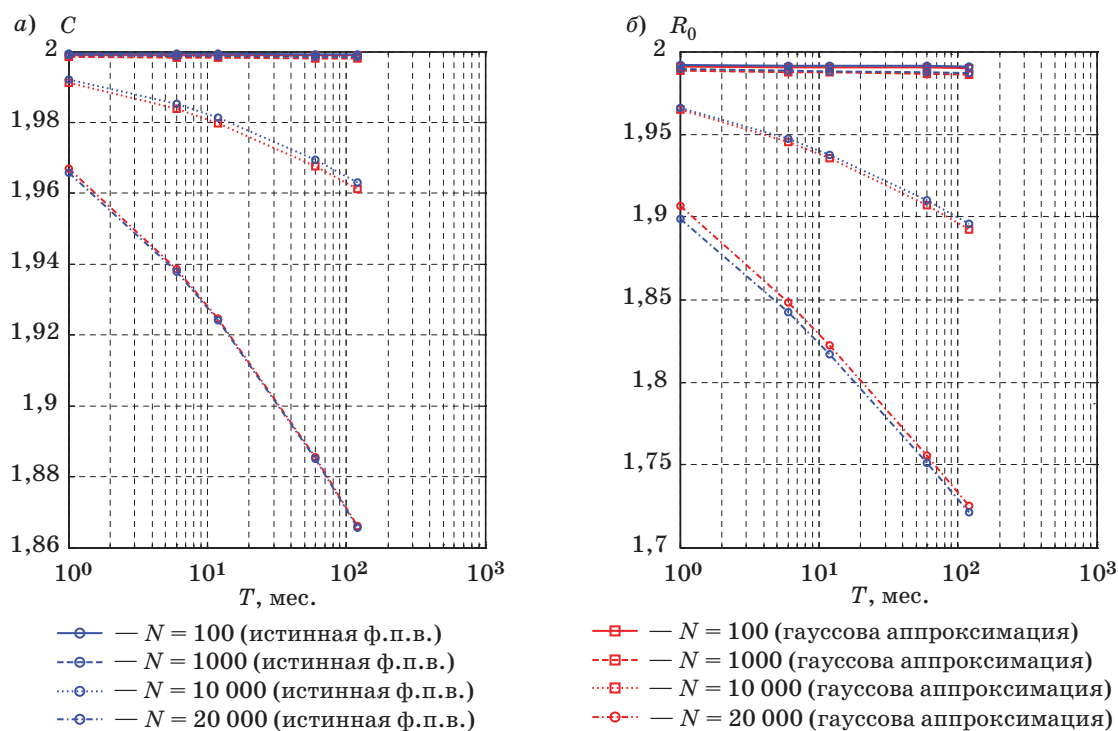
Рисунок 5, а и б представляет зависимость величин C и R_0 от времени хранения данных T при различных значениях числа циклов перезаписи N , выполненных за это время. При увеличении времени хранения T параметры канала записи ухудшаются, и величины C и R_0 , характеризующие потенциально достижимую скорость кодирования, уменьшаются. Это уменьшение оказывается крайне небольшим при $N \leq 10^3$. При $N \geq 10^4$



■ Рис. 3. Зависимости пропускной способности C (а) и вычислительной скорости R_0 (б) от числа циклов перезаписи N



■ Рис. 4. Оптимальные распределения $p_x(\cdot)$ в зависимости от числа циклов перезаписи N



■ Рис. 5. Зависимости величин C (а) и R_0 (б) от времени хранения данных T

и увеличении времени хранения данных T до года и более снижение становится значительнее, хотя и остается в пределах 5...8 %. Как и ранее, характеристики, полученные исходя из истинных плотностей вероятности $p_{y|x}(\cdot)$, задающих канал, и гауссовых приближений к этим плотностям, оказываются очень близкими.

Заключение

В работе были рассмотрены теоретически допустимые пределы скорости кодирования, для которых возможно построение надежных систем хранения данных для одной из имеющихся моделей flash-памяти. Эта модель описывает систему с независимыми многоуровневыми ячейками.

Выходное значение представляет собой случайную величину, определенную на множестве одиночных элементов памяти, распределение которой существенно зависит от ряда мешающих факторов, включающих влияние циклов записи/стирания, взаимную интерференцию ячеек и влияние времени хранения данных. В рассматриваемой модели предполагается отсутствие группирования искажений, поэтому канал хранения данных описывается моделью без памяти с непрерывным выходом. Такая модель задает наилучшее использование выхода канала. Для этих условий были построены границы случайного кодирования в форме, которая допускает получение численных результатов. Наряду с «истинными» функциями плотности вероятности были рассмо-

трены гауссовы приближения к ним. Эти приближения оказались в ряде случаев достаточно точными, что позволило получить точные оценки вычислительной скорости R_0^* в явном виде как функции от параметров гауссовых аппроксимаций.

На основе представленных результатов можно сформулировать следующие выводы:

— для многих практически важных случаев значения пропускной способности и вычислительной скорости оказываются близкими (их разница составляет величину от долей процента до нескольких процентов);

— пропускная способность и вычислительная скорость принимают значения, немного уступающие максимальной для этой модели величине 2 бит (для ряда случаев, представляющих интерес, разница лежит в пределах 5–10 %);

— в большинстве рассмотренных случаев оптимальное распределение на входе канала, доставляющее максимум величинам R_0 и C , оказывается очень близким к равномерному распределению, что может быть аргументом в пользу рассмотрения корректирующих кодов без ограничений на использование различных символов входного алфавита;

— пропускная способность канала записи убывает сравнительно медленно с ростом числа циклов перезаписи до величины порядка нескольких тысяч; после значения порядка 10^4 убывание становится более заметным;

— длительность хранения данных снижает достижимую скорость кодирования, но при небольшом числе циклов перезаписи ($N \leq 10^3$) это уменьшение оказывается незначительным; при $N \geq 10^4$ и увеличении времени хранения данных T до года и более снижение становится более значительным, хотя и остается в пределах 5–8 %;

— на основании полученных границ случайного кодирования можно указать ориентировочные пределы изменения основных параметров блочных кодов, которые оказываются полезными для поиска значений параметров реальных систем кодирования для каналов записи/считывания данных.

Работа выполнена при поддержке Минобрнауки РФ при проведении научно-исследовательской работы в рамках проектной части государственного задания в сфере научной деятельности по заданию № 2.2716.2014/К от 17.07.2014.

Литература

1. Nonvolatile Memory Technologies with Emphasis on Flash: A Comprehensive Guide to Understanding and Using Flash Memory Devices/ J. E. Brewer, M. Jill, editors. — Wiley-IEEE Press, 2008. — 792 p.
2. Bez R., Camerlenghi E., Modelli A., Visconti A. Introduction to Flash Memory// Proc. of the IEEE. 2003. Vol. 91. N 4. P. 489–502.
3. Yan Li, et al. A 16 Gb 3b/cell NAND Flash Memory in 56 nm with 8 MB/s Write Rate/Li Yan, Lee Seungpil, Fong Yupin, Pan Feng, Kuo Tien-Chien, Park Jong, T. Samaddar, Nguyen Hao, Mui Man, Htoo Khin, T. Kamei, M. Higashitani, E. Yero, Kwon Gyuwan, P. Kliza, Wan Jun, T. Kaneko, H. Maejima, H. Shiga, M. Hamada, N. Fujita, K. Kanebako, E. Tarn, A. Koh, I. Lu, C. Kuo, Pham Trung, J. Huynh, Nguyen Qui, H. Chibvongodze, M. Watanabe, K. Oowada, G. Shah, Woo Byungki, Gao Ray, J. Chan, J. Lan, P. Hong, Peng Liping, D. Das, D. Ghosh, V. Kalluru, S. Kulkarni, R. Cernea, S. Huynh, D. Pantelakis, Wang Chi-Ming, K. Quader // Proc. of the IEEE Intern. Solid-State Conf. (ISSCC/8). Feb. 3–7, 2008. P. 506–507.
4. Yan Li, et al. 128 Gb 3b/cell NAND Flash Memory in 19 nm with 18 MB/s Write Rate and 400 MB/s Toggle Mode / Li Yan, Lee Seungpil, K. Oowada, Nguyen Hao, Qui Nguyen, N. Mokhlesi, C. Hsu, J. Li, V. Ramachandra, T. Kamei, M. Higashitani, T. Pham, M. Honma, Y. Watanabe, K. Ino, Le Binh, Woo Byungki, Htoo Khin, Tseng Tai-Yuan, L. Pham, F. Tsai, Kim Kwang-ho, Chen Yi-Chieh, She Min, Yuh Jong, A. Chu, Chen Chen, R. Puri, Lin Hung-Szu, Chen Yi-Fang, W. Mak, J. Huynh, Chan Jim, M. Watanabe, D. Yang, G. Shah, P. Souriraj, D. Tadepalli, S. Tenu-gu, Gao Ray, V. Popuri, B. Azarbayjani, R. Madpur, J. Lan, E. Yero, Pan Feng, P. Hong, Yong Kang Jang, F. Moogat, Fong Yupin, R. Cernea, S. Huynh, C. Trinh, M. Mofidi, R. Shrivastava, K. Quader // Proc. of the IEEE Intern. Solid-State Conf. (ISSCC/12). San Francisco, CA, USA, Feb. 19–23, 2012. P. 436–437.
5. Shibata N., et al. A 70 nm 16 Gb 16-level NAND Flash Memory / N. Shibata, H. Maejima, K. Isobe, K. Iwasa, M. Fujiu, M. Nakagawa, T. Shimizu, M. Honma, S. Hoshi, T. Kawaai, K. Kanebako, S. Yoshikawa, H. Tabata, A. Inoue, T. Takahashi, T. Shano, Y. Komatsu, K. Nagaba, M. Kosakai, N. Motohashi, K. Kanazawa, K. Imamiya, Nakai H. Nakai, M. Lasser, M. Murin, A. Meir, A. Eyal, M. Shlick // IEEE Journal on Solid-State Circuits. 2008. Vol. 43. N 4. P. 929–937.
6. Yaakobi E., et al. Error Correcting Coding for Flash Memories / E. Yaakobi, J. Ma, A. Caulfield, L. Grupp, S. Swanson, P. H. Siegel, J. K. Wolf // CMRR Report. San Diego: University of California; Center for Magnetic Recording Research. Winter 2009. N 31. P. 6.
7. Yaakobi E., et al. Error Characterization and Coding Schemes for Flash Memories/ E. Yaakobi, J. Ma, A. Caulfield, L. Grupp, S. Swanson, P. H. Siegel, J. K. Wolf // IEEE Global Communications Workshop on Application of Communication Theory to Emerging Memory Technologies (ACTEMT 2010). Miami, FL, USA, Dec. 6–10, 2010. P. 1856–1860.

8. Wang X., Dong G., Pan L., Zhou R. Error Correction Codes and Signal Processing in Flash Memory. Ch. 3 in "Flash Memories"/ Prof. Igor Stievano (Ed.). 2011. <http://www.intechopen.com/books/flash-memories/error-correction-codes-and-signalprocessing-in-flash-memory> (дата обращения: 26.10.2014).
9. Dong G., et al. Estimating Information-Theoretical NAND Flash Memory Storage Capacity and its Implication to Memory System Design Space Exploration / G. Dong, Y. Pan, N. Xie, C. Varanasi, T. Zhang // IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2012. Vol. 20. N 9. P. 1705–1714.
10. Huang X., et al. Multilevel Flash Memories: Channel modeling, Capacities and Optimal Coding Rates / X. Huang, A. Kavcic, X. Ma, G. Dong, T. Zhang // International Journal on Advances in Systems and Measurement. 2013. Vol. 6. N 3&4. P. 364–373. http://www.iariajournals.org/systems_and_measurements/sysmea_v6_n34_2013_paged.pdf (дата обращения: 26.10.2014).
11. Sun F., Rose K., Zhang T. On the Use of Strong BCH Codes for Improving Multilevel NAND Flash Memory Storage Capacity. http://www.researchgate.net/publication/254376882_On_the_Use_of_Strong_BCH_Codes_for_Improving_Multilevel_NAND_Flash_Memory_Storage_Capacity (дата обращения: 26.10.2014).
12. Gallager R. G. Information Theory and Reliable Communication. — N. Y.: John Wiley & Sons, 1968. — 588 p.
13. Viterbi A. J., Omura J. K. Principles of Digital Communication and Coding. — N. Y.: McGraw-Hill, 1979. — 560 p.
14. Wozenkraft J. M., Jacobs I. M. Principles of Communication Engineering. — N. Y.: John Wiley & Sons, 1968. — 720 p.

UDC 621.391

doi:10.15217/issn1684-8853.2016.2.56

Information Theory Analysis of Multilevel Flash Memory. Part 1: Channel Model and Random Coding BoundsTrofimov A. N.^a, PhD, Tech., Associate Professor, andrei.trofimov@vu.spb.ruTaubin F. A.^a, Dr. Sc. Tech., Professor, ftaubin@yahoo.com^aSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., Saint-Petersburg, 190000, Russian Federation

Introduction: Flash memory is one of the fastest growing segments in the global semiconductor industry. Due to the steadily increasing recording density, high read/write speed, low power consumption and long service life, flash memory is used to store data in a very wide field of applications. Increasing the recording density achieved by decreasing the physical size of the memory cell, and the increasing number of states of a memory cell lead to a decrease in data storage reliability, which requires the use of error-correcting coding. **Purpose:** We study theoretically achievable values for the main parameters of error-correcting coding in a multilevel NAND flash memory model, depending on the recording/storage conditions. **Results:** Theoretically achievable coding rate limits have been obtained within which it is possible to build reliable memory for one of possible flash memory models treating a memory page as a system with independent multilevel cells. Random coding bounds have been derived for this model, and some numerical results presented. Using these bounds, we establish certain trade-off relationships between the achievable coding rate and the key memory parameters such as the number of program/erase cycles and the data retention time. Gaussian approximation for the channel model has also been considered. This approximation in some cases is sufficiently accurate to get an exact expression for the bound on the channel cutoff rate R_0^* as a function of the Gaussian approximation parameters. It is shown that in many cases the channel capacity and the channel cutoff rate are very close. The difference between them is less or about several percent. **Practical relevance:** The obtained results allow us to accurately estimate the reduction of the achievable recording rate (bit/cell) in a multilevel NAND flash memory caused by the growing number of the program/erase cycles and the longer data retention time. This gives us the possibility to specify the bounds for the program/erase cycles and data retention time which would ensure the required recording rate.

Keywords — Multilevel NAND Flash Memory, Recording Channel, Coding Theorem, Channel Capacity, Random Coding Bound.

References

1. Brewer J. E., Jill M., editors. *Nonvolatile Memory Technologies with Emphasis on Flash: A Comprehensive Guide to Understanding and Using Flash Memory Devices*. Wiley-IEEE Press, 2008. 792 p.
2. Bez R., Camerlenghi E., Modelli A., Visconti A. Introduction to Flash Memory. *Proc. of the IEEE*, 2003, vol. 91, no. 4, pp. 489–502.
3. Yan Li, Seungpil Lee, Yupin Fong, Feng Pan, Tien-Chien Kuo, Jong Park, Samaddar T., Hao Nguyen, Man Mui, Khin Htoo, Kamei T., Higashitani M., Yero E., Gyuwan Kwon, Kliza P., Jun Wan, Kaneko T., Maejima H., Shiga H., Hamada M., Fujita N., Kanabako K., Tarn E., Koh A., Lu I., Kuo C., Trung Pham, Huynh J., Qui Nguyen, Chibvongodze H., Watanabe M., Oowada K., Shah G., Byungki Woo, Ray Gao, Chan J., Lan J., Hong P., Liping Peng, Das D., Ghosh D., Kaluru V., Kulkarni S., Cernea R., Huynh S., Pantelakis D., Chi-Ming Wang, Quader K. A 16 Gb 3b/cell NAND Flash Memory in 56 nm with 8 MB/s Write Rate. *Proc. of the IEEE Intern. Solid-State Conf. (ISSCC)*, 3–7 Feb. 2008, pp. 506–507.
4. Yan Li, Seungpil Lee, Oowada K., Hao Nguyen, Qui Nguyen, Mokhlesi N., Hsu C., Li J., Ramachandra V., Kamei T., Higashitani M., Pham T., Honma M., Watanabe Y., Ino K., Binh Le, Byungki Woo, Khin Htoo, Tai-Yuan Tseng, Pham L., Tsai F., Kwang-ho Kim, Yi-Chieh Chen, Min She, Jong Yuh, Chu A., Chen Chen, Puri R., Hung-Szu Lin, Yi-Fang Chen, Mak W., Huynh J., Jim Chan, Watanabe M., Yang D., Shah G., Souriraj P., Tadepalli D., Tenugu, S., Ray Gao, Popuri V., Azarbayjani B., Madpur R., Lan J., Yero E., Feng Pan, Hong P., Jang Yong Kang, Moogat F., Yupin Fong, Cernea R., Huynh S., Trinh C., Mofidi M., Shrivastava R., Quader K. 128 Gb 3b/cell NAND Flash Memory in 19 nm with 18 MB/s Write Rate and 400 MB/s Toggle Mode. *Proc. of the IEEE Intern. Solid-State Conf. (ISSCC/12)*, San Francisco, CA, USA, 2012, pp. 436–437.
5. Shibata N., Maejima H., Isobe K., Iwasa K., Nakagawa M., Fujii M., Shimizu T., Honma M., Hoshi S., Kawaai T., Kanabako K., Yoshikawa S., Tabata H., Inoue A., Takahashi T., Shano T., Komatsu Y., Nagaba K., Kosakai M., Motohashi N.,

- Kanazawa K., Imamiya K., Nakai H., Lasser M., Murin M., Meir A., Eyal A., Shlick M. A 70 nm 16 Gb 16-level NAND Flash Memory. *IEEE Journal on Solid-State Circuits*, 2008, vol. 43, no. 4, pp. 929–937.
6. Yaakobi E., Ma J., Caulfield A., Grupp L., Swanson S., Siegel P. H., Wolf J. K. Error Correcting Coding for Flash Memories. *CMRR Report*, Winter 2009, Number 31, University of California, Center for Magnetic Recording Research, San Diego, 2009, p. 6.
 7. Yaakobi E., Ma J., Caulfield A., Grupp L., Swanson S., Siegel P. H., Wolf J. K. Error Characterization and Coding Schemes for Flash Memories. *IEEE Global Communications Workshop on Application of Communication Theory to Emerging Memory Technologies (ACTEMT 2010)*, Miami, FL, USA, 6–10 Dec. 2010, pp. 1856–1860.
 8. Wang X., Dong G., Pan L., Zhou R. *Error Correction Codes and Signal Processing in Flash Memory*. Ch. 3 in “Flash Memories”, Prof. Igor Stievano (Ed.), 2011. InTech. Available at: <http://www.intechopen.com/books/flash-memories/error-correction-codes-and-signalprocessing-in-flash-memory> (accessed 26 October 2014).
 9. Dong G., Pan Y., Xie N., Varanasi C., Zhang T. Estimating Information-Theoretical NAND Flash Memory Storage Capacity and its Implication to Memory System Design Space Exploration. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2012, vol. 20, no. 9, pp. 1705–1714.
 10. Huang X., Kavcic A., Ma X., Dong G., Zhang T. Multilevel Flash Memories: Channel Modeling, Capacities and Optimal Coding Rates. *Intern. Journal on Advances in Systems and Measurement*, 2013, vol. 6, no. 3&4, pp. 364–373. Available at: http://www.iariajournals.org/systems_and_measurements/sysmea_v6_n34_2013_paged.pdf (accessed 26 October 2014).
 11. Sun F., Rose K., Zhang T. *On the Use of Strong BCH Codes for Improving Multilevel NAND Flash Memory Storage Capacity*. Available at: http://www.researchgate.net/publication/254376882_On_the_Use_of_Strong_BCH_Codes_for_Improving_Multilevel_NAND_Flash_Memory_Storage_Capacity (accessed 26 October 2014).
 12. Gallager R. G. *Information Theory and Reliable Communication*. New York, John Wiley & Sons, 1968. 588 p.
 13. Viterbi A. J., Omura J. K. *Principles of Digital Communication and Coding*. New York, McGraw-Hill, 1979. 560 p.
 14. Wozenkraft J. M., Jacobs I. M. *Principles of Communication Engineering*. New York, John Wiley & Sons, 1968. 720 p.

Уважаемые авторы!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

Статьи должны содержать изложение новых научных результатов. Название статьи должно быть кратким, но информативным. В названии недопустимо использование сокращений, кроме самых общепринятых (РАН, РФ, САПР и т. п.).

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала Word шрифтом Times New Roman размером 13, поля не менее двух сантиметров.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание (при отсутствии — должность), полное название организации, аннотация и ключевые слова на русском и английском языках, электронные адреса авторов, которые по требованию ВАК должны быть опубликованы на страницах журнала. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы.

Статьи авторов, не имеющих ученой степени, рекомендуется публиковать в соавторстве с научным руководителем, наличие подписи научного руководителя на рукописи обязательно; в случае самостоятельной публикации обязательно предоставляйте заверенную по месту работы рекомендацию научного руководителя с указанием его фамилии, имени, отчества, места работы, должности, ученого звания, ученой степени — эта информация будет опубликована в ссылке на первой странице.

Формулы набирайте в Word, не используя формульный редактор (Mathtype или Equation), при необходимости можно использовать формульный редактор; для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта никогда не пользуйтесь вкладкой Other..., используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; в формулах не отделяйте пробелами знаки: + = -.

Для набора формул в Word никогда не используйте Конструктор (на верхней панели: «Работа с формулами» — «Конструктор»), так как этот ресурс предназначен только для внутреннего использования в Word и не поддерживается программами, предназначенными для изготовления оригинал-макета журнала.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Иллюстрации в текст не завершаются и предоставляются отдельными файлами, поддающимися редактированию:

— рисунки, графики, диаграммы, блок-схемы предоставляются в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio 4, 5, 2002–2003 (*.vsd); Coreldraw (*.cdr); Excel (*.xls); Word (*.doc); AdobeIllustrator (*.ai); AutoCad (*.dxf); Matlab (*.ps, *.pdf или экспорт в формат *.ai);

— если редактор, в котором Вы изготавливаете рисунок, не позволяет сохранить в векторном формате, используйте функцию экспорта (только по отношению к исходному рисунку), например, в формат *.ai, *.esp, *.wmf, *.emf, *.svg;

— фото и растровые — в формате *.tif, *.png с максимальным разрешением (не менее 300 pixels/inch).

Наличие подписуточных подписей обязательно (желательно не повторяющих дословно комментарии к рисункам в тексте статьи).

В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, e-mail), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате *.tif, *.png с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40×55 мм;

— экспертное заключение.

Список литературы составляется по порядку ссылок в тексте и оформляется следующим образом:

— для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц;

— для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Список литературы оформляйте двумя отдельными блоками по образцам lit.dot на сайте журнала (<http://i-us.ru/paperrules>) по разным стандартам: Литература — СИБИД РФ, References — один из мировых стандартов.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Оформление статей».

Контакты

Куда: 190000, Санкт-Петербург,
Б. Морская ул., д. 67, ГУАП, РИЦ
Кому: Редакция журнала «Информационно-управляющие системы»
Тел.: (812) 494-70-02
Эл. почта: ius.spb@gmail.com
Сайт: www.i-us.ru

МОДЕЛИ КОДИРОВАНИЯ ПОЛИМОДАЛЬНОЙ ИНФОРМАЦИИ

О. О. Басов^а, канд. техн. наук

И. С. Кипяткова^б, канд. техн. наук, старший научный сотрудник

А. И. Савельев^б, научный сотрудник

И. А. Саитов^а, доктор техн. наук, профессор

^аАкадемия Федеральной службы охраны Российской Федерации, Орел, РФ

^бСанкт-Петербургский институт информатики и автоматизации РАН, Санкт-Петербург, РФ

Постановка проблемы: имеющиеся предпосылки для отказа от традиционных принципов разделения передаваемой информации на услуги связи и возможность реализации полимодальных инфокоммуникационных систем обуславливают необходимость разработки теоретических основ построения последних. **Цель:** формализация моделей кодирования полимодальной информации, а также обоснование методов ее передачи на основе результатов моделирования. **Методы:** для оценки качества кодирования многомодальной информации при заданных ресурсах сети передачи данных предложена модификация теоретико-множественной модели. **Результаты:** представлены аналитические выражения, описывающие зависимости качества кодирования полимодальной информации и объема, требуемого для передачи максимального числа сообщений различных модальностей, от параметров модифицированной модели. На основе анализа существующих и перспективных методов передачи сделан вывод о том, что современные средства телекоммуникаций способны обеспечить доставку полимодальной информации с требуемым качеством к абонентскому терминалу корреспондента. Перспективной в этом смысле является реализация сети передачи данных, инвариантной к применяемой технологии передачи.

Ключевые слова — полимодальная инфокоммуникационная система, модальность, абонентский терминал, сеть передачи данных, теоретико-множественная модель.

Введение

В процессе совершенствования национальные информационные инфраструктуры развитых стран эволюционировали [1] от узкоспециализированных сетей связи (телеграфных, телефонных и др.) сначала в телекоммуникационные сети, а затем в инфокоммуникационные системы. В настоящее время уже имеются предпосылки для отказа от традиционных принципов разделения передаваемой информации на услуги связи и возможность реализации полимодальных инфокоммуникационных систем (ПИКС) [2–7]. Под ПИКС следует понимать взаимоувязанную совокупность систем обработки и хранения информации и их объединяющих телекоммуникационных сетей. Целью ее функционирования является сбор, обработка, хранение, защита, передача и распределение, отображение и использование многомодальной информации, учитывающей смысл передаваемых сообщений, личность абонентов (пользователей), их настроение, физиологическое и психоэмоциональное состояния [8–10].

Очевидно, что новые возможности инфокоммуникационного обслуживания пользователей ПИКС должны быть заложены на этапе проектирования. При синтезе (моделировании, проектировании, оптимизации) ПИКС ее удобно представить в виде двух условно автономных подсистем: совокупности абонентских терминалов и сети передачи данных (СПД) [11]. Общее модельное понятие «сеть передачи данных» используется для составных физических или логических кана-

лов связи и формирующих их узлов коммутации транспортной сети и сетей доступа. При ориентации на предоставление ПИКС единственной услуги в виде «соединение с сетью» практически важными являются две основные задачи синтеза.

Прямая. Имеются сведения об объемах сообщений (наборе модальностей), которые необходимо передать. Требуется определить объем ресурсов СПД, чтобы обеспечить заданное качество приема переданной полимодальной информации.

Двойственная. Имеются заданные ресурсы СПД. Требуется передать максимальный объем сообщений заданного качества (в том числе с оптимальным выбором модальностей, методов их обработки: объединения/разделения, синхронизации, кодирования).

Решение прямой и двойственной задач синтеза ПИКС требует разработки теоретических конструкций (моделей) для оценки качества кодирования многомодальной информации при заданных ресурсах СПД и объема указанного ресурса, необходимого для передачи максимального числа сообщений различных модальностей с заданным качеством, а также обоснования выбора методов передачи таких сообщений на основе результатов моделирования.

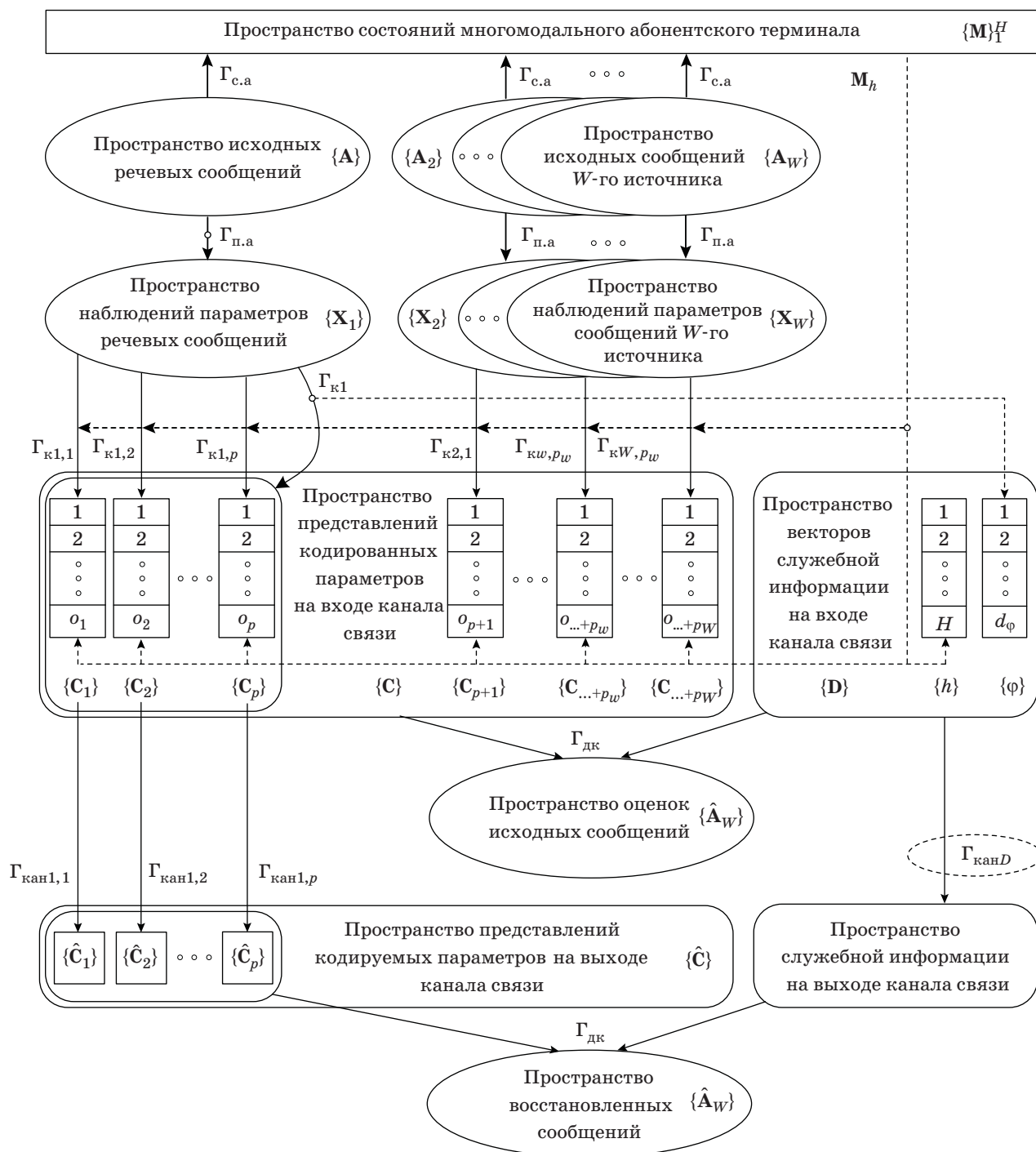
Математические модели кодирования сообщений различных модальностей

Практика математического описания систем кодирования, использующих процедуры параметрического анализа и синтеза, скалярного и век-

торного квантования, свидетельствует о целесообразности применения для этих целей аппарата теории множеств и теории вероятностей [12–15].

С учетом существующей модели адаптивной системы кодирования речевого сигнала [16, 17] теоретико-множественная модель подсистемы кодирования сообщений различных модальностей может быть представлена следующим образом (рисунок).

Модель основана на предположении, что коммуникативный акт абонентов происходит с обязательным использованием вербального речевого канала. При этом используется известное представление сигналов на входе подсистемы кодирования в виде сообщений дискретного источника, рассматриваемых как точки в P -мерном пространстве [12, 18]. Каждую точку этого пространства удобно представить в виде вектора $\mathbf{A}_P \in \{\mathbf{R}^P\}$.



■ Теоретико-множественная модель кодирования сообщений различных модальностей

Индекс P , показывающий размерность вектора сообщения, для сокращения записи в дальнейших рассуждениях опущен. Пространство, образуемое множеством точек $\mathbf{A}$, рассматривается как пространство сообщений $\{\mathbf{A}\}$. При этом в дальнейшем рассматривается параметрическое кодирование избыточных сообщений [16].

Входом подсистемы кодирования (см. рисунок) являются множества $\{\mathbf{A}_w\}$, $w = \overline{1, W}$ реализаций сообщений источников различных модальностей. К ее внутренним параметрам относятся:

- 1) число W источников (модальностей);
- 2) количество K_w , $w = \overline{1, W}$ значений случайной величины, описывающей w -й источник сообщений на интервале анализа T_a ;
- 3) вид отображений параметрического анализа $\Gamma_{п.а}$, определяемый вариантом формирования, комбинирования и отображения множества параметров $\mathbf{X}_w$ анализируемых сообщений $\mathbf{A}_w$;
- 4) число p_w , $w = \overline{1, W}$ и вид представления кодируемых параметров $\mathbf{X}_{w,j}$, $j = \overline{1, p_w}$ источников многомодальной информации;
- 5) вид отображений $\Gamma_{к\ w,j}$, $w = \overline{1, W}$, $j = \overline{1, p_w}$, определяющих процедуры квантования наблюдаемых параметров источников многомодальной информации;
- 6) число уровней квантования o_j каждого ($j = \overline{1, p_1 + \dots + p_W}$) наблюдаемого параметра источников информации, определяющих мощности подмножеств кодированных параметров $\{\mathbf{C}_j\}$ на входе СПД;
- 7) вид отображения статистического анализа $\Gamma_{с.а}$, определяющего процедуру классификации исходных сообщений W источников;
- 8) число H состояний абонентского терминала $\{\mathbf{M}\}$, определяющих множество вариантов распределения информационной емкости канала связи по множествам кодируемых параметров W источников;

$$\{\mathbf{M}\} = \{No_{\Gamma_{к1,1}}\} \times \dots \times \{No_{\Gamma_{к1,p}}\} \times \dots \times \{No_{\Gamma_{кW,1}}\} \times \dots \\ \times \{No_{\Gamma_{кW,p_W}}\} \times \{o_1\} \times \dots \times \{o_p\} \times \dots \\ \times \{o_{p_1 + \dots + p_{W-1} + 1}\} \times \dots \times \{o_{p_1 + \dots + p_W}\},$$

где $No_{\Gamma_{к\ w,j}}$ — номер возможного отображения $\Gamma_{к\ w,j}$, $w = \overline{1, W}$, $j = \overline{1, p_1 + \dots + p_W}$. Значения $o_j = 1$ обеспечивают возможность исключения j -х подпространств из структуры пространства кодируемых параметров. В частном случае однозначного соответствия значений o_j и отображений $\Gamma_{к\ w,j}$ множество

$$\{\mathbf{M}\} = \{o_1\} \times \dots \times \{o_p\} \times \dots \\ \times \{o_{p_1 + \dots + p_{W-1} + 1}\} \times \dots \times \{o_{p_1 + \dots + p_W}\}.$$

Для высокоадаптивной системы [10] число возможных состояний абонентского терминала может быть бесконечно большим: $H = |\{\mathbf{M}\}| = \infty$.

С учетом искажений, получаемых при кодировании, и влияния канала связи, задаваемого отображением $\Gamma_{кан}$, при известном отображении $\Gamma_{дек}$, однозначно определяющем процедуру декодирования, математическое описание всей подсистемы кодирования относительно внешнего параметра, характеризующего качество кодирования сообщений различных модальностей, может быть представлено в следующем виде:

$$D_{сум} = \left(\sum_{w=1}^W \frac{D_w[\{\mathbf{A}\}, \{\hat{\mathbf{A}}\}]}{P_{C_w}} + \right. \\ \left. + \sigma \sum_{q=1}^W \left(\frac{D_w[\{\mathbf{A}\}, \{\hat{\mathbf{A}}\}]}{P_{C_w}} - \frac{D_q[\{\mathbf{A}\}, \{\hat{\mathbf{A}}\}]}{P_{C_q}} \right)^2 \right), \quad (1)$$

где $D_w[\{\mathbf{A}\}, \{\hat{\mathbf{A}}\}]$, $w = \overline{1, W}$ — среднеквадратическая ошибка между множествами исходных и восстановленных сообщений или энергия шума при восстановлении сообщений w -го источника; σ — эмпирический коэффициент, определяющий

степень влияния штрафа; $P_{C_w} = \sum_{i=1}^U \mathbf{A}_{wi}^T \mathbf{A}_{wi}$ — энергия сообщений w -го источника.

Представленный выше перечень внутренних параметров подсистемы кодирования сообщений различных модальностей позволяет сформулировать математическую модель такой системы относительно внешнего параметра — скорости передачи (выдачи информации):

$$B_{(W)} = \frac{\log \prod_{w=1}^W \prod_{j=p_{w-1}+1}^{p_w} o_j + r + \log d_\phi + \log H}{T_a} \text{ [бит/с]}, \quad (2)$$

где $r = \left\lceil \log_2 \left(1 + \sum_{i=1}^{t_n} C_n^i \right) \right\rceil$ — число проверочных разрядов, а t_n — исправляющая способность помехоустойчивого кода; C_n^i — число сочетаний из n по i ; $d_\phi = |\{\phi\}|$ — мощность подпространства $\{\phi\} \in \{\mathbf{D}\}$, характеризующего множество вариантов помехоустойчивого кодирования.

Для подсистемы кодирования с одним источником информации (речевым сообщением) [16] скорость передачи

$$B_{(1)} = \frac{\log \prod_{j=1}^p o_j + r + \log d_\phi}{T_a} \text{ [бит/с]}. \quad (3)$$

В соответствии с теорией информации [18, 19] наибольший выигрыш относительно суммарной скорости передачи может быть достигнут при совместном кодировании зависимых источников,

т. е. источников, между которыми имеются корреляционные взаимосвязи. При этом величина выигрыша совместного кодирования по сравнению с раздельным определяется степенью корреляции источников.

В работе [12] теоретически обоснованы условия, при которых возникают корреляционные зависимости между сообщениями независимых избыточных источников речевых и видеосообщений на ограниченных интервалах наблюдения. Наличие данных зависимостей подтверждено экспериментально путем постановки и решения задачи поиска коэффициентов линейного преобразования сообщений независимых дискретных источников. Наличие корреляционных зависимостей между другими модальностями требует проведения соответствующих исследований. Показано, что выигрыш по величине (1) при совместном кодировании по сравнению с раздельным увеличивается при уменьшении скорости передачи сообщений одного источника. При одинаковом количестве данных, затрачиваемых на кодирование каждого сообщения W источников, величина ошибки при их восстановлении в точке приема при совместном кодировании существенно меньше, чем при раздельном, причем выигрыш систем совместного кодирования по ошибке аппроксимации каждого сообщения в сравнении с системами раздельного кодирования растет при увеличении числа кодируемых источников (модальностей). В указанном случае в большей степени расширяется номенклатура скоростей B_j .

Возможность реализации последней в существующих и перспективных СПД обсуждается [7], основные выводы представлены ниже.

Методы передачи многомодальной информации через сеть передачи данных

Самой доступной сегодня для ПИКС является электрически инвариантная СПД с коммутацией пакетов. В таких системах единица канального ресурса составит значение порядка 64 Кбит/с, что обеспечит достаточную чувствительность всей системы к изменению числа активных модальностей. Для моделирования таких СПД ПИКС имеется возможность использовать существующий инструментарий анализа современных мультипротокольных телекоммуникационных сетей, в качестве источников нагрузки которых будут выступать многомодальные абонентские терминалы. Для построения теоретической модели СПД с коммутацией пакетов необходимо выделить, определить и формально описать ее структурные элементы, охарактеризовать их основное назначение и специфику процессов функционирования.

Основными элементами сетей связи с накоплением являются абонентские терминалы, узлы коммутации и связывающие их каналы связи. Существенное влияние на качество синхронизации модальностей при их автономной обработке и передаче соответствующих блоков данных через СПД с коммутацией пакетов будет оказывать сетевая задержка. Под сетевой задержкой понимается интервал времени между моментом, когда первый бит базы данных поступает в сеть с коммутацией пакетов, и моментом, когда последний бит этого блока покидает сеть.

Наличие сетевой задержки и ее случайный характер свидетельствуют о том, что в СПД с коммутацией пакетов целесообразно реализовать совместную обработку модальностей в абонентском терминале, а соответствующие базы данных передавать с использованием механизмов обеспечения гарантированного качества обслуживания в СПД.

Перспективным для ПИКС является реализация в СПД принципов построения оптических транспортных сетей, инвариантных к технологии передачи. Метод передачи в таких СПД будет определяться текущим множеством активных пользователей и составом используемых модальностей в их абонентских терминалах.

В настоящее время вместо традиционных гомогенных волоконно-оптических линий связи на практике все чаще реализуются гетерогенные волоконно-оптические линии связи, использующие в различных спектральных каналах оптического волокна разные сетевые технологии. При наличии средств согласования режимов функционирования многомодальных абонентских терминалов и сети передачи данных появляется возможность организации в последней каналов и трактов, адаптивно «подбирающих» наиболее эффективную сетевую технологию (*SyTDM* или *StTDM*) под текущее распределение объемов и типов модальностей в сообщениях.

Изложенное выше подтверждает необходимость редукции объекта исследования (ПИКС) на две условно-автономные подсистемы: множества многомодальных абонентских терминалов и СПД, способную с заданным качеством обеспечить обслуживание поступающей нагрузки. Естественно, такой подход обуславливает необходимость оптимизации в дальнейшем целого ряда функциональных характеристик элементов СПД для обслуживания многомодальных сообщений от абонентского терминала, что потребует новых теоретических исследований в предметной области.

Заключение

Возможность обслуживания сообщений от многомодальных абонентских терминалов узлами коммутации и каналами связи осложняется

расширившейся номенклатурой скоростей передачи. Анализ существующих и перспективных методов передачи блоков данных в СПД позволил сделать вывод о том, что современные средства телекоммуникаций способны обеспечить доставку объектов учета с требуемым качеством к абонентскому терминалу корреспондента. В СПД с *SyTDM* потребуется значительный расход ресурса пропускной способности, однако в такой инфраструктуре проще обеспечить синхронизацию модальностей. В СПД с *StTDM* потребуется использовать более сложные алгоритмы обеспечения гарантированного качества доставки

блоков данных. Из-за наличия стохастических задержек блоков данных в таких СПД сложнее обеспечить синхронизацию модальностей при их автономной обработке. В настоящее время имеются объективные предпосылки для реализации СПД, инвариантной к применяемой технологии передачи. Эта возможность определяется широким (от абонента к абоненту) применением оптических технологий в транспортной инфраструктуре ПИКС.

Исследование выполнено при частичной поддержке бюджетной темы № 0073-2014-0005 и грантов РФФИ № 15-07-06744 и 16-19-00044.

Литература

1. **Саитов И. А.** Основы теории построения защищенных мультипротокольных оптических транспортных сетей телекоммуникационных систем: монография. — Орел: Академия ФСО России, 2008. — 220 с.
2. **Басов О. О., Саитов И. А.** Основные каналы межличностной коммуникации и их проекция на инфокоммуникационные системы // Тр. СПИИРАН. 2013. № 7. С. 122–140.
3. **Gregory F. D., Dai L.** Multisensory Information Processing for Enhanced Human-Machine Symbiosis // Human Interface and the Management of Information. Information and Knowledge Design: Proc. of 17th Intern. Conf., HCI Intern. 2015, Los Angeles, CA, USA, Aug. 2–7, 2015/ Yamamoto (Ed.). 2015. Part I. LNCS 9172. P. 354–365.
4. **Ронжин А. Л., Карпов А. А.** Многомодальные интерфейсы: основные принципы и когнитивные аспекты // Тр. СПИИРАН. 2006. № 3. Т. 1. С. 300–319.
5. **Карпов А. А., Лале А., Ронжин А. Л.** Многомодальные ассистивные системы для интеллектуального жилого пространства // Тр. СПИИРАН. 2011. № 19. С. 48–64.
6. **Карпов А. А., Ронжин А. Л.** Многомодальные интерфейсы в автоматизированных системах управления // Изв. вузов. Приборостроение. 2005. Т. 48. № 7. С. 9–14.
7. **Басов О. О., Саитов И. А.** Методы передачи полимодальной информации // Научно-технический вестник информационных технологий, механики и оптики. 2015. Т. 15. № 2. С. 293–299.
8. **White T. L., Krausman A. S.** Effects of Inter-Stimulus Interval and Intensity on the Perceived Urgency of Tactile Patterns. Applied Ergonomics. 2015. Vol. 48. P. 121–129.
9. **Ронжин Ал. Л., Будков В. Ю., Ронжин Ан. Л.** Формирование профиля пользователя на основе аудиовизуального анализа ситуации в интеллектуальном зале совещаний // Тр. СПИИРАН. 2012. № 4(23). С. 482–494.
10. **Басов О. О., Карпов А. А., Саитов И. А.** Методологические основы синтеза полимодальных инфокоммуникационных систем государственного управления: монография. — Орел: Академия ФСО России, 2015. — 272 с.
11. **Басов О. О., Саитов И. А.** Качество функционирования и эффективность полимодальных инфокоммуникационных систем // Тр. СПИИРАН. 2014. № 1(32). С. 152–170.
12. **Устинов А. А.** Стохастическое кодирование видео- и речевой информации: в 2 ч. / под ред. проф. В. Ф. Комаровича. — СПб.: ВАС, 2005.
13. **Speech Recognition and Coding. New Advances and Trends. NATO ASI Series.** — Germany, 1995. — 464 p.
14. **Липкин И. А.** Статистическая радиотехника. Теория информации и кодирования. — М.: Вузовская книга, 2002. — 216 с.
15. **Афанасьев А. А., Богачев Г. В., Рыболовлев А. А.** Теоретико-множественная модель липредера с адаптацией распределения информационных ресурсов к статистическим характеристикам кодируемых параметров. — Орел: Академия ФАПСи, 2002. — 60 с.
16. **Басов О. О.** Математическая модель системы кодирования речевого сигнала с многопараметрической адаптацией // Телекоммуникации. 2008. № 7. С. 7–13.
17. **Basov O. O.** A Conceptual Model of Multicriterion Adaptation of the Linear Predictive Voice Coding Procedure // Telecommunications and Radio Engineering. 2009. N 10(68). P. 923–931.
18. **Csiszar I., Körner Ya.** Information Theory: Coding Theorems for Discrete Memoryless Systems. — Cambridge University Press, 2011. — 600 p.
19. **Feng H., Effros M.** On the Rate Loss of Multiple Description Source Codes // IEEE Trans. on Information Theory. 2005. Vol. 51. N 2. P. 671–683.

UDC 62-503.57:621.391

doi:10.15217/issn1684-8853.2016.2.68

Polymodal Information Encoding Models

Basov O. O.^a, PhD, Tech., oobasov@mail.ruKipyatkova I. S.^b, PhD, Tech., Senior Researcher, kipyatkova@ias.spb.suSaveliev A. I.^b, Researcher, saveliev@ias.spb.suSaitov I. A.^a, Dr. Sc., Tech., Professor, akramovish@mail.ru^aAcademy of Federal Agency of Protection of Russian Federation, 35, Priborostroitel'naya St., 302034, Orel, Russian Federation^bSaint-Petersburg Institute for Informatics and Automation of RAS, 39, 14 Line, V. O., 199178, Saint-Petersburg, Russian Federation

Introduction: The existing prerequisites to reject traditional principles of dividing transmitted information into communication services and to implement polymodal infocommunication systems (PICS) illustrate the necessity to develop a theoretical basis for building PICS. **Purpose:** Formalization of models for encoding polymodal information, and justification of methods of its transmission based on simulation results. **Methods:** For assessing the quality of multimodal information encoding with given data transmission network resources, we propose a modification of a set-theoretical model. **Results:** Analytical expressions are presented which describe the dependency of the multimodal information encoding quality, as well as that of the volume required for transferring the maximum number of messages of different modalities, on the modified model parameters. Based on the analysis of the existing and prospective transmission methods, a conclusion can be made that modern telecommunication facilities can ensure the delivery of multimodal information with the required quality to the subscriber terminal. In this respect, a data network implementation invariant to the applied transmission technology has a great potential.

Keywords — Polymodal Infocommunication System, Modality, Subscriber Terminal, Data Transmission Network, Set-Theoretical Model.

References

1. Saitov I. A. *Osnovy teorii postroeniia zashchishchennykh mul'tiprotokol'nykh opticheskikh transportnykh setei telekommunikatsionnykh sistem* [Fundamentals of the Theory of Building Secure Multiprotocol Optical Transport Networks of Telecommunication Systems]. Orel, Akademiia FSO Rossii, 2008. 220 p. (In Russian).
2. Basov O. O., Saitov I. A. Basic Channels of Interpersonal Communication and their Projection on the Infocommunications Systems. *Trudy SPIIRAN*, 2013, no. 7(30), pp. 122–140 (In Russian).
3. Gregory F. D., Dai L. Multisensory Information Processing for Enhanced Human-Machine Symbiosis. *Proc. of 17th Intern. Conf. "Human Interface and the Management of Information. Information and Knowledge Design"*, Los Angeles, CA, USA, 2015, part I, LNCS 9172, pp. 354–365.
4. Ronzhin A. L., Karpov A. A. Multimodal Interfaces: Main Principles and Cognitive Aspects. *Trudy SPIIRAN*, 2006, iss. 3, vol. 1, pp. 300–319 (In Russian).
5. Karpov A. A., Lale A., Ronzhin A. L. Multimodal Assistive Systems for a Smart Living Environment. *Trudy SPIIRAN*, 2011, no. 19, pp. 48–64 (In Russian).
6. Karpov A. A., Ronzhin A. L. Multimodal Interfaces in Automated Control Systems. *Izvestiia vuzov. Priborostroenie*, 2005, no. 7(48), pp. 9–14 (In Russian).
7. Basov O. O., Saitov I. A. Methods of Transfer of Polymodal Information. *Nauchno-tekhnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki*, 2015, no. 2(15), pp. 293–299 (In Russian).
8. White T. L., Krausman A. S. Effects of Inter-Stimulus Interval and Intensity on the Perceived Urgency of Tactile Patterns. *Applied Ergonomics*, 2015, vol. 48, pp. 121–129.
9. Ronzhin A. L., Budkov V. Yu., Ronzhin A. L. User Profile Forming Based on Audiovisual Situation Analysis in Smart Meeting Room. *Trudy SPIIRAN*, 2012, no. 4(23), pp. 482–494 (In Russian).
10. Basov O. O., Karpov A. A., Saitov I. A. *Metodologicheskoe osnovy sinteza polimodal'nykh infokommunikatsionnykh sistem gosudarstvennogo upravleniia* [Methodological Basis for Synthesis of Polymodal Infocommunication Systems for the State Administration]. Orel, Akademiia FSO Rossii Publ., 2015. 271 p. (In Russian).
11. Basov O. O., Saitov V. A. Functioning Quality and Effectiveness of Polymodal Infocommunicational Systems. *Trudy SPIIRAN*, 2014, no. 1(32), pp. 152–170 (In Russian).
12. Ustinov A. A. *Stokhasticheskoe kodirovanie video- i rechevoi informatsii* [Stochastic Coding of Video and Speech Information]. V. F. Komarov ed. Saint-Petersburg, VAS Publ., 2005. Vol. 1. (In Russian).
13. Speech Recognition and Coding. New Advances and Trends. NATO ASI Series. Germany, 1995. 464 p.
14. Lipkin I. A. *Statisticheskaya radiotekhnika. Teoriia informatsii i kodirovaniia* [Statistical radio Engineering. Information Theory and Coding]. Moscow, Vuzovskaia kniga Publ., 2002. 216 p. (In Russian).
15. Afanas'ev A. A., Bogachev G. V., Rybolovlev A. A. *Teoretiko-mnozhestvennaia model' lipredera s adaptatsiei raspredeleniia informatsionnykh resursov k statisticheskim kharakteristikam kodiruemykh parametrov* [A Set-Theoretical Model of a Linear Prediction Vocoder with Adaptation of Information Resources Distribution to the Statistical Characteristics of Encoded Parameters]. Orel, Akademiia FAPSI Publ., 2002. 60 p. (In Russian).
16. Basov O. O. Mathematic Model of the Speech Coding System with a Multiparameter Adaptation. *Telekommunikatsii*, 2008, no. 7, pp. 7–13 (In Russian).
17. Basov O. O. A Conceptual Model of Multicriterion Adaptation of the Linear Predictive Voice Coding Procedure. *Telecommunications and Radio Engineering*, 2009, no. 10(68), pp. 923–931.
18. Csiszar I., Körner Ya. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011. 600 p.
19. Feng H., Effros M. On the Rate Loss of Multiple Description Source Codes. *IEEE Trans. on Information Theory*, 2005, no. 2(51), pp. 671–683.

ПЕРЕДАЧА ГИПЕРСПЕКТРАЛЬНЫХ ВИДЕОДАНЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ПО РАДИОКАНАЛАМ С ОГРАНИЧЕННОЙ ПРОПУСКНОЙ СПОСОБНОСТЬЮ

Г. Н. Мальцев^а, доктор техн. наук, профессор

И. А. Козин^а, канд. техн. наук, доцент

^аВоенно-космическая академия им. А. Ф. Можайского, Санкт-Петербург, РФ

Постановка проблемы: использование гиперспектральной съемки одновременно в сотнях спектральных каналов является устойчивой тенденцией развития космических систем дистанционного зондирования Земли. Оно приводит к опережающему росту информативности бортовых оптико-электронных систем наблюдения по отношению к пропускной способности радиоканалов передачи информации с космических аппаратов. В результате возникает проблемная ситуация, состоящая в том, что современные радиотехнические системы передачи информации с ограниченной пропускной способностью не позволяют передавать полные объемы зарегистрированных гиперспектральных видеоданных как при непосредственной оперативной передаче, так и при воспроизведении из бортового запоминающего устройства. Отсюда следует необходимость разработки технологий сокращения (сжатия) потока гиперспектральных видеоданных для их передачи на наземные пункты приема информации с использованием спутниковых радиотехнических систем передачи информации с ограниченной пропускной способностью. **Результаты:** анализ информационной производительности существующих бортовых гиперспектральных оптико-электронных систем показал, что она достигает единиц гигабит в секунду и в перспективе будет увеличиваться, при этом пропускная способность радиоканалов передачи видеоданных с космических аппаратов с учетом возможностей используемых диапазонов частот и ограничений технической реализации не превышает нескольких сотен мегабит в секунду. Для сокращения объемов передаваемой информации обосновано совместное использование программных алгоритмов сжатия видеоданных в каждом спектральном канале и выбора для передачи ограниченного числа наиболее информативных спектральных каналов из всех зарегистрированных гиперспектральных видеоданных. Это обеспечивает сокращение объема передаваемой информации в несколько десятков раз и согласование информационной производительности оптико-электронных систем наблюдения с пропускной способностью радиоканалов передачи видеоданных. Определены достижимые коэффициенты сжатия гиперспектральных видеоданных при совместном использовании различных программных методов сжатия и методов выбора наиболее информативных спектральных каналов. **Практическая значимость:** полученные результаты могут быть использованы при проектировании и применении космических систем дистанционного зондирования Земли с гиперспектральной съемкой для выбора технологий и режимов передачи с космических аппаратов данных наблюдения по радиоканалам с ограниченной пропускной способностью.

Ключевые слова — дистанционное зондирование Земли, гиперспектрометр, спектральный канал, скорость передачи информации, сокращение избыточности, сжатие видеоданных.

Введение

Современные космические системы дистанционного зондирования Земли (ДЗЗ) характеризуются большим разнообразием, в их составе используются космические аппараты (КА), которые ведут съемку в различных диапазонах электромагнитного спектра с разрешением от нескольких километров до нескольких десятков сантиметров [1–5]. В настоящее время одной из ведущих тенденций развития космических систем ДЗЗ является переход к широкому использованию бортовых многоспектральных и гиперспектральных оптико-электронных систем (ОЭС), работающих в видимом и ИК-диапазонах и осуществляющих съемку в сотнях спектральных каналов [4–7]. Гиперспектральные ОЭС наблюдения являются результатом эволюционного развития многоспектральных систем. Принятое деление ОЭС наблюдения на многоспектральные и гиперспектральные основывается на числе ка-

налов регистрации видеоданных: у многоспектральных систем — десятки, у гиперспектральных систем — сотни спектральных каналов.

Гиперспектральная аппаратура наблюдения — гиперспектрометры; они представляют собой инновационный тип приборов ДЗЗ для исследования природных ресурсов, охраны окружающей среды, предупреждения и ликвидации последствий природных катастроф и техногенных аварий, решения задач метеорологии, климатологии и многих других областей [4, 5]. При этом создание гиперспектральных ОЭС отражает одну из тенденций развития информационных систем и технологий, связанную с опережающим увеличением производительности систем сбора информации по сравнению с увеличением пропускной способности систем передачи информации. Несмотря на то что в многоспектральных и гиперспектральных ОЭС возможен «обмен» пространственного разрешения на спектральное, объемы данных наблюдения, которые

необходимо передавать с КА, ведущих многоспектральную и гиперспектральную съемку, резко возрастают, и остро встает проблема оперативной передачи гиперспектральных данных по радиоканалам с ограниченной пропускной способностью. В настоящей статье приводятся результаты оценки информационной производительности КА ДЗЗ с бортовыми гиперспектральными ОЭС и рассматриваются возможности уменьшения объемов передаваемых гиперспектральных видеоданных за счет сжатия и выбора для передачи ограниченного числа наиболее информативных спектральных каналов.

Анализ информативности гиперспектральных данных космических систем ДЗЗ

Широкие возможности ДЗЗ с использованием гиперспектральных ОЭС открываются благодаря совместному анализу пространственных (топологических) и спектральных характеристик наблюдаемых объектов, процессов и явлений. При этом используется «гиперкуб» данных, определяющий трехмерную матрицу пространственного (по двум координатам) и спектрального (по третьей координате) распределения яркостей регистрируемого излучения. Это создает условия для обнаружения и распознавания по топологическим и спектральным признакам различных объектов атмосферы, океана, суши, геологической среды, оценки их физико-химических свойств и определения других значимых характеристик. Однако данные наблюдения необходимо передавать на наземные пункты приема информации (ППИ), в ряде случаев — оперативно, в масштабе времени, близком к реальному, и существенную роль приобретает соотношение между информационной производительностью аппаратуры ДЗЗ и пропускной способностью используемых радиоканалов передачи информации.

При рассмотрении информационных характеристик бортовых гиперспектральных ОЭС и систем передачи с КА видеоданных будем понимать под ДЗЗ наблюдение из космоса как земной поверхности и находящихся на ней объектов, так и состояния атмосферы, хотя с точки зрения прикладных задач и типа космической системы эти задачи соответствуют космическим системам ДЗЗ (в их обычном понимании) и метеорологии [1, 3]. Современными тенденциями в разработке бортовых гиперспектральных ОЭС являются возрастание детальности космических снимков, увеличение полос обзора и захвата, увеличение числа спектральных каналов гиперспектрометров [2–4]. Все это приводит к увеличению объема «гиперкуба» и информационной производительности КА ДЗЗ с бортовыми гиперспектральными ОЭС, ко-

торая начинает существенно превышать пропускную способность существующих спутниковых радиотехнических систем передачи информации с КА. Поэтому при проектировании и создании перспективных космических систем ДЗЗ с гиперспектральной съемкой необходимо уделять серьезное внимание вопросам передачи данных наблюдения с КА и выбору соответствующей технологии передачи видеоданных, учитывающей соотношение между производительностью источника передаваемой информации (гиперспектрометра) и достижимыми в имеющихся системах передачи информации скоростями передачи.

Рассмотрим характеристики информационной производительности гиперспектральных систем ДЗЗ. При кадровой съемке количество информации в одном кадре видеоданных (сцене), регистрируемых бортовой гиперспектральной ОЭС, определяется выражением [8]

$$I = L_1 L_2 g b / \Delta l^2, \quad (1)$$

где L_1, L_2 — размеры сцены вдоль и поперек трассы; g — число разрядов для цифровой записи зарегистрированной яркости в элементе матричного фотоприемника (радиометрическое разрешение); b — число регистрируемых спектральных каналов; Δl — линейное разрешение на местности, соответствующее элементу (пикселю) матричного фотоприемника.

Расчеты по формуле (1) объемов регистрируемых видеоданных для современных гиперспектральных аэрокосмических систем ДЗЗ, приведенных в табл. 1, показывают, что объем инфор-

■ Таблица 1. Объемы видеоданных, регистрируемых современными бортовыми гиперспектрометрами

Аппаратура / носитель	L_1 , км	L_2 , км	l , м	g , бит	b	I , Гбит
MODIS Terra / Aqua	2500	2330	250; 500; 1000	12	36	5,3
Hiperion / EO-1	100	7800	30	16	220	2,84
Artemis / TacSAT-3	100	7800	30	16	400	5,17
AIRS / Aqua	1650	1650	13500	12	2378	0,4
CHRIS / Proba-1	14	14	18; 36	18	19; 63	0,36
ГС / МКА-ФКИ	35	35	52	12	150	0,76
ГСА / «Ресурс-П»	25	25	25	14	96–255	1,07–2
«Лептон» / Авиационный	175	175	0,35	12	156	0,436

мации, формируемой при гиперспектральной съемке в пределах одного кадра, может составлять от 0,5 до 5 Гбайт. А суммарный суточный поток данных ДЗЗ, формируемых при комплексном аэрокосмическом мониторинге с использованием многоспектральных и гиперспектральных ОЭС различного назначения, оценивается величиной до 0,5 Тбайт и более [4, 8].

Наиболее жесткие требования к пропускной способности радиоканалов передачи информации предъявляются в режиме непосредственной передачи — при передаче данных ДЗЗ в масштабе времени, близком к реальному. В этом случае скорость передачи информации в радиоканале должна соответствовать информационной производительности аппаратуры наблюдения. Для получающего наиболее широкое распространение в бортовых ОЭС КА ДЗЗ метода маршрутной съемки с автоматической разверткой в направлении полета КА (метод «заметания» — push broom) [2, 4] скорость передачи информации, требуемая для цифровой передачи видеоданных от бортовой гиперспектральной ОЭС в реальном масштабе времени, определяется выражением [9]

$$v = V_{КА} S g b / \Delta l^2, \quad (2)$$

где $V_{КА}$ — скорость КА для заданной высоты полета $H_{КА}$; S — ширина полосы захвата; g — число разрядов при цифровой передаче зарегистрированной яркости в элементе матричного фотоприемника. Полагается, что размер элемента матричного фотоприемника согласован с линейным разрешением на местности.

При использовании метода push broom регистрация видеоданных выполняется при строчном сканировании линейки матричного фотоприемника, а развертка в направлении полета КА осуществляется за счет его собственного движения. Отношение $N = S / \Delta l$ дает число элементов разрешения бортовой ОЭС в полосе обзора. Если полагать, что скорость КА $V_{КА}$ соответствует первой космической скорости для круговой орбиты высотой $H_{КА}$, то выражение (2) для скорости цифрового потока видеоданных, поступающих от бортовой ОЭС наблюдения, может быть преобразовано к виду

$$v = \frac{N g b}{\Delta l} \sqrt{\frac{\mu}{R_3 + H_{КА}}}, \quad (3)$$

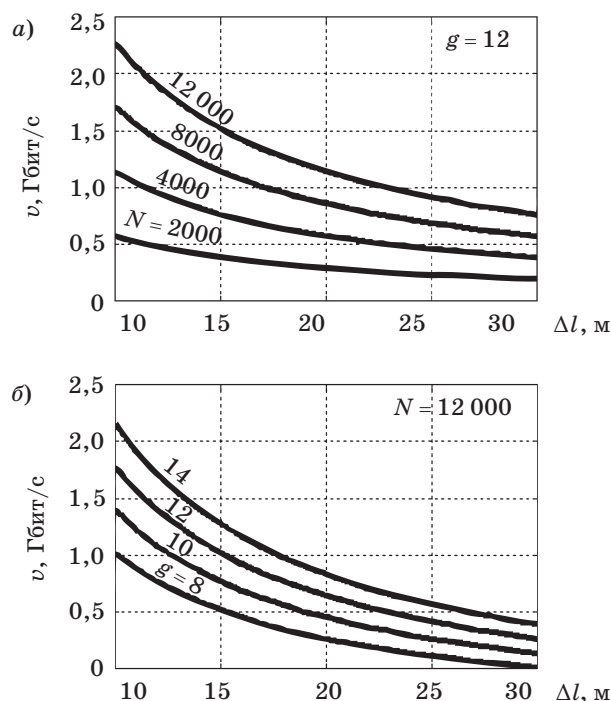
где $\mu = 3,986 \cdot 10^5 \text{ км}^3 \cdot \text{с}^{-2}$ — гравитационная постоянная; $R_3 = 6387 \text{ км}$ — радиус Земли.

Выполнены расчеты по формуле (3) скорости потока видеоданных v от бортовой гиперспектральной ОЭС наблюдения в зависимости от линейного разрешения на местности Δl для $H_{КА} = 600 \text{ км}$, $b = 220$ при различных значениях

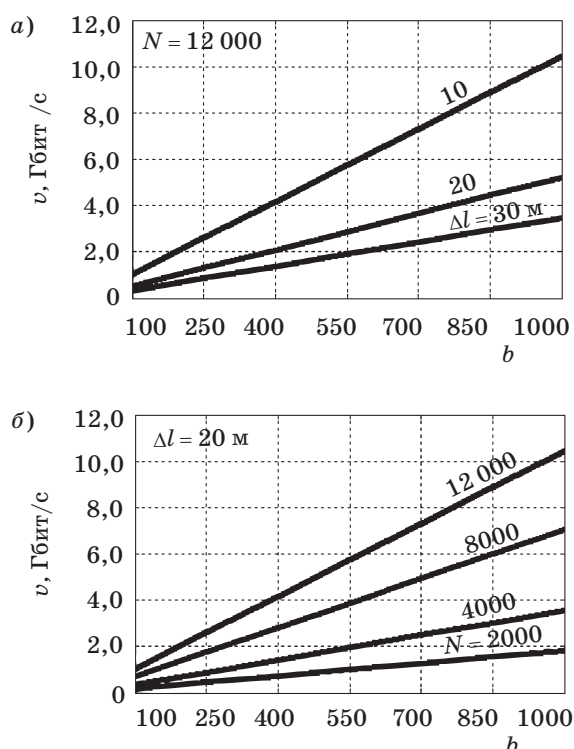
числа элементов разрешения бортовой ОЭС N в полосе обзора S (рис. 1, а) и числа разрядов при цифровой передаче видеоданных g (рис. 1, б).

Расчеты по формуле (3) скорости потока видеоданных v от бортовой гиперспектральной ОЭС наблюдения в зависимости от числа спектральных каналов b для $H_{КА} = 600 \text{ км}$, $g = 12$ произведены при различных значениях линейного разрешения на местности Δl (рис. 2, а) и различных значениях числа элементов разрешения бортовой ОЭС N в полосе обзора S (рис. 2, б).

При анализе информационных возможностей космических гиперспектральных систем ДЗЗ практический интерес представляет диапазон значений параметров Δl от десятков до сотен метров, b — до нескольких сотен и N — до десятка тысяч, что соответствует системам среднего пространственного разрешения с широкой полосой обзора и числу каналов существующих гиперспектрометров, разработанных для использования на КА ДЗЗ. Приведенные результаты расчетов показывают, что при таких характеристиках бортовой ОЭС наблюдения в режиме непосредственной передачи потока гиперспектральных видеоданных необходимо обеспечить скорости передачи информации в сотни мегабит в секунду. С уменьшением Δl и увеличением b требуемая скорость передачи данных в реальном масштабе



■ Рис. 1. Скорость поступления потока видеоданных v в зависимости от линейного разрешения на местности Δl при различных значениях числа элементов разрешения в полосе обзора N (а) и числа разрядов при цифровой передаче видеоданных g (б)



■ **Рис. 2.** Скорость поступления потока видеоданных v в зависимости от числа спектральных каналов b при различных значениях линейного разрешения на местности Δl (а) и числа элементов разрешения в полосе обзора N (б)

времени будет составлять единицы гигабит в секунду. Так, например, чтобы передать в режиме реального времени с КА, находящегося на орбите высотой $H_{КА} = 600$ км, гиперспектральные видеоданные с параметрами $b = 500$, $\Delta l = 20$ м, $g = 16$ и при $N = 4000$ потребуются обеспечить скорость передачи не менее 12 Гбит/с.

Расчетные характеристики информационной производительности современных КА ДЗЗ с гиперспектральной съемкой приведены в табл. 2. Анализ оценок информационной производительности показывает, что для представленных в таблице КА ДЗЗ передача видеоданных с КА на наземные ППИ в масштабе времени, близком к реальному, требует обеспечения скорости передачи информации в сотни мегабит–единицы гигабит в секунду. Только при таких скоростях передачи информации происходит согласование скорости поступления информации от бортовой аппаратуры наблюдения со скоростью ее цифровой передачи по каналу связи. Высокие требования к скорости передачи информации могут также иметь место и при воспроизведении из бортового запоминающего устройства больших объемов записанных видеоданных при ограниченном времени прохождения КА в зоне видимости наземного ППИ.

■ **Таблица 2.** Информационная производительность современных КА ДЗЗ с гиперспектральной съемкой

Наименование КА / государство (оператор)	Число каналов / диапазон съемки, мкм	Разрешение на местности, м	Полоса обзора, км	Информационная производительность датчика, Мбит/с
MightySat II.1. (Sindri) / США	256 / 0,47÷10,05	30–250; 26–51	20	430
Earth Observing-1 (EO-1) / США	220 / 0,4÷2,5	30	7,7	240
EOS AM-1 (Terra) / США	14 / 0,52÷11,6	15; 30; 90	60	85
EOS PM-1 (Aqua) / США	36 / 0,41÷14,39	250; 500; 1000	2300	10
Proba / США	150 / 0,4÷10,5	5; 18; 36	14	1000
OFEC-7 / Израиль	200 / 0,35÷3	10–15	11	2000
«Ресурс-П» / Россия	255 / 0,4÷1,1	25–30	25–30	500
«Кондор-Э» / Россия (проект)	256 / 0,5÷0,8	20	15×12	725
«Астрогон-Вулкан» / Россия (проект)	700 / 0,2÷1,4	5	15×12	7000

Анализ характеристик радиотехнических систем передачи данных ДЗЗ

Рассмотрим возможности существующих спутниковых радиотехнических систем передачи информации по передаче видеоданных гиперспектрального ДЗЗ на наземные ППИ. В настоящее время для передачи данных с КА ДЗЗ используются радиотехнические системы передачи информации, работающие в различных диапазонах частот — от ОВЧ/УВЧ-диапазонов (135–150 МГц, 400–470 МГц) до Х-диапазона (7450–8400 МГц) [10, 11]. Практический интерес представляет анализ возможностей высокоскоростной передачи гиперспектральных видеоданных в Х-диапазоне, который является в настоящее время основным для цифровой передачи с КА ДЗЗ больших объемов видеоданных со скоростью до 320 Мбит/с в одном радиоканале. Для сравнения: в S-диапазоне, также широко используемом для передачи с КА данных ДЗЗ, скорость передачи не выше 15 Мбит/с в одном радиоканале. В табл. 3

представлены значения скоростей передачи информации в некоторых современных спутниковых радиотехнических системах передачи видеоданных с КА ДЗЗ, работающих в X-диапазоне, и используемых в них методов модуляции (структур сигналов) [10].

В перспективе для высокоскоростной передачи с КА данных ДЗЗ планируется использование Ка-диапазона (23–28 ГГц), однако в настоящее время спутниковые радиотехнические системы передачи информации, работающие в этом диапазоне, находятся в стадии разработки [11]. При этом от перехода в Ка-диапазон не следует ожидать значительного увеличения скоростей передачи информации свыше сотен мегабит в секунду. В перспективных лазерных высокоскоростных системах передачи информации с КА на первом этапе их развития также рассматривается возможность реализации скоростей передачи информации в сотни мегабит в секунду (до 600 Мбит/с) [12]. Это обусловлено тем, что реализуемые скорости передачи информации определяются не только потенциальными возможностями диапазона частот по увеличению ширины полосы радиоканала, но и быстродействием электронных устройств формирования сигналов, кодирования и декодирования информации. Существенную роль играют также ограничения технической реализации, связанные, в том числе, с используемыми структурами сигналов и методами кодирования при обеспечении требуемого энергетического потенциала высокоскоростной радиолинии передачи с КА видеоданных.

■ **Таблица 3.** Характеристики радиолиний передачи видеоданных с КА ДЗЗ

КА	Частотный диапазон радиолинии, МГц	Скорость передачи информации, Мбит/с	Метод модуляции
Terra, Aqua	8160÷8212,5	13; 150	QPSK, SQPSK
EO-1, Cartosat-1 (IRS-P5)	8125÷8300	52,5	QPSK
«Ресурс-П», QuickBird-2, Ikonos, Orbview-3,5	8025÷8400	75; 150; 320; 370 (2 канала)	BPSK, QPSK, OQPSK, SQPSK
«Метеор-3М», «Аркон 2», БелКА	8025÷8400	15; 30; 61; 123	BPSK, QPSK
Pleiades	7450÷8400	600 (2 канала)	8PSK
Eros C (A, B)	7450÷8400	70; 280; 455	QPSK
Landsat-7, ALOS	8105÷8342,5	150; 138,76	QPSK, AQPSK

Обеспечение высокой скорости передачи информации в радиолинии «борт–земля» также зависит от характеристик приемной аппаратуры наземного ППИ, поскольку именно ее характеристиками в большинстве случаев обеспечивается требуемый энергетический потенциал радиолинии передачи информации с КА. Требуемые характеристики радиолинии передачи с КА данных ДЗЗ при скоростях передачи информации 5–15 Мбит/с обеспечиваются при выходной мощности бортового радиопередатчика до 10 Вт с использованием слабонаправленной бортовой передающей спиральной антенны размером 50–60 см и наземной приемной зеркальной антенны диаметром 0,6–2,0 м, а для обеспечения скоростей передачи информации 150–300 Мбит/с при тех же характеристиках бортовых передатчиков и антенн необходимы наземные приемные антенны диаметром 5–9 м. В первом случае наземный ППИ может быть реализован как в мобильном исполнении, так и в стационарном варианте, во втором случае — только в стационарном варианте. При этом дальнейшее увеличение скорости передачи информации осуществляется за счет передачи данных одновременно по нескольким высокоскоростным каналам. Использование в качестве наземных ППИ малогабаритных мобильных приемных станций с малыми диаметрами приемной зеркальной антенны приводит к необходимости увеличения размеров и направленности бортовых антенн и, как следствие, усложнения аппаратуры их наведения. В противном случае обеспечить высокую скорость передачи с КА данных ДЗЗ невозможно.

Современные отечественные комплексы приема информации с КА ДЗЗ обеспечивают в X-диапазоне скорость приема информации до 150 Мбит/с при диаметре наземной антенны 2,4 м, скорость приема информации до 320 Мбит/с — при диаметре наземной антенны 5–9 м [11, 13], при этом в ряде систем для достижения более высоких скоростей передачи информации используется параллельная передача по нескольким радиоканалам.

Таким образом, в существующих и перспективных спутниковых системах передачи видеоданных с КА ДЗЗ скорости передачи информации составляют сотни мегабит в секунду. Практически следует ориентироваться на типовую скорость передачи информации 300 Мбит/с для X-диапазона с возможностью ее увеличения в перспективе до 600 Мбит/с–1,2 Гбит/с при параллельной многоканальной передаче, переходе в Ка-диапазон и использовании лазерных систем. Однако эти возможности систем передачи информации уже сейчас не согласуются с информационными возможностями многих гиперспектральных систем ДЗЗ, и в перспективе это рассогласование будет только

усиливаться. Поэтому необходим поиск путей согласования информационной производительности аппаратуры ДЗЗ с пропускной способностью радиоканалов передачи данных. Это становится возможным при сокращении избыточности зарегистрированных гиперспектральных данных наблюдения и снижении на этой основе объемов передаваемых с КА ДЗЗ видеоданных.

Современные подходы к сокращению избыточности гиперспектральных видеоданных

Проблему сокращения избыточности гиперспектральных видеоданных при их передаче с КА ДЗЗ следует решать по двум направлениям. Первое направление соответствует традиционному подходу к сокращению избыточности видеoinформации — их сжатию с использованием различных алгоритмов сжатия видеоданных [14, 15]. Второе направление является характерным для обработки данных многоспектральных и гиперспектральных ОЭС наблюдения и связано с выбором для оперативной передачи и обработки наиболее информативных спектральных каналов [16, 17].

Применение различных алгоритмов сжатия видеоданных получило в настоящее время широкое распространение как для передачи изображений, так и, даже в большей степени, для их хранения. Уменьшение требуемой скорости передачи информации при использовании алгоритмов сжатия видеоданных определяется коэффициентом сжатия используемого алгоритма. Существуют алгоритмы сжатия с потерями и без потерь. Наиболее известными и универсальными алгоритмами сжатия изображений без потерь являются программные алгоритмы WinRar, WinZip, алгоритм Lossless JPEG на основе стандарта сжатия JPEG. Известны и находят применение также другие программные алгоритмы сжатия видеоданных, такие как Adaptive Binary Optimization (ABO), GIF, Progressive Graphics File (PGF), PNG, Qbit Lossless Codec, TIFF, WMPhoto, RunLengthEncoding (RLE), LZW [14, 15].

Учитывая необходимость последующей тематической обработки данных наблюдения и распознавания наблюдаемых объектов, необходимо использовать алгоритмы сжатия видеоданных ДЗЗ без потерь, не допускающие искажений статистических яркостных характеристик восстановления данных. Однако коэффициент сжатия для алгоритмов сжатия без потерь ниже, чем для алгоритмов сжатия с потерями. При сжатии отдельных изображений, что в случае гиперспектральных видеоданных соответствует сжатию каждого изображения зарегистрированного «гиперкуба» в отдельности, коэффициент сжатия

составляет единицы. При совместном сжатии нескольких изображений, зарегистрированных в нескольких спектральных каналах, коэффициент сжатия может быть увеличен в 1,5–2 раза при учете межканальной корреляции [4, 15].

Для обеспечения высоких значений коэффициентов сжатия видеоданных необходимо использовать алгоритмы сжатия с потерями информации, способные обеспечить коэффициенты сжатия порядка десятков и более. В основе методов сжатия изображений с потерями лежат алгоритмы фрактального и вейвлет-сжатия, алгоритмы Low JPEG2000 и High JPEG2000 на основе стандарта сжатия JPEG, методы сжатия на основе дифференциального анализа и др. [14]. В случае их использования для сокращения избыточности видеоданных ДЗЗ необходимо следить, чтобы потери информации не приводили к ухудшению качества решения задач тематической обработки данных наблюдения.

Оценки коэффициента сжатия $K_{сж}$ для 10 различных программных алгоритмов сжатия гиперспектральных данных представлены на рис. 3. В качестве тестового изображения использовался тестовый гиперспектральный снимок размером 1200×4000 элементов, с числом спектральных каналов $b = 500$, радиометрическим разрешением $g = 16$, объемом 4,8 Гбайт. Такой объем видеоданных без сжатия может быть передан по радиоканалу со скоростью передачи информации 100–150 Мбит/с за 4–5 мин, что соизмеримо с длительностью сеанса связи с низкоорбитальным КА ДЗЗ при его прохождении в зоне радиовидимости наземного ППИ. Алгоритмы сжатия без потерь 1–8 обеспечивают уменьшение объема передаваемых видеоданных до 2,5 раза, а алгоритмы сжатия с потерями 9–10 — уменьшение объема передаваемых видеоданных до 5,5 раза.

Недостатком рассмотренных алгоритмов сжатия изображений, не позволяющим достичь



■ Рис. 3. Оценки коэффициента сжатия $K_{сж}$ для различных алгоритмов сжатия видеоданных: 1 — WinRar; 2 — WinZip; 3 — TIFF; 4 — PNG; 5 — RLE; 6 — LZW; 7 — JPEG; 8 — JPEG2000 LS; 9 — JPEG2000 Low; 10 — JPEG2000 High

высоких коэффициентов сжатия многоспектральных и гиперспектральных данных, является то, что в них для сжатия используются различные переходы в цветоразностном пространстве, что плохо работает в случае совместного сжатия изображений, зарегистрированных в различных спектральных каналах. Кроме того, в большинстве алгоритмов сжатия изображений, адаптированных для сжатия многоспектральных и гиперспектральных видеоданных, многомерные спектральные данные рассматриваются как набор независимых двумерных изображений в пространственных координатах, а взаимосвязь данных по третьей — спектральной — координате не учитывается.

Учет корреляции и взаимосвязи гиперспектральных данных по спектральной координате учитывается при выборе для передачи и обработки наиболее информативных спектральных каналов. Анализ показывает, что имеет место значительная корреляция спектральных характеристик большинства природных и техногенных объектов и типов подстилающей поверхности в оптическом и ИК-диапазонах по длине волны. Это позволяет существенным образом сократить объем анализируемых гиперспектральных данных за счет сокращения числа передаваемых и поступающих в обработку видеоданных различных спектральных каналов. Данные, зарегистрированные в соседних спектральных каналах, сильно коррелированы между собой, и для решения большинства прикладных задач ДЗЗ полный объем многоспектральных и гиперспектральных данных избыточен и может быть сокращен. При этом число спектральных каналов, поступающих в обработку, может быть сокращено до 6–20, что соответствует величине существенной спектральной размерности (числу значимых спектральных составляющих) многоспектральных данных оптических наблюдений [16, 18].

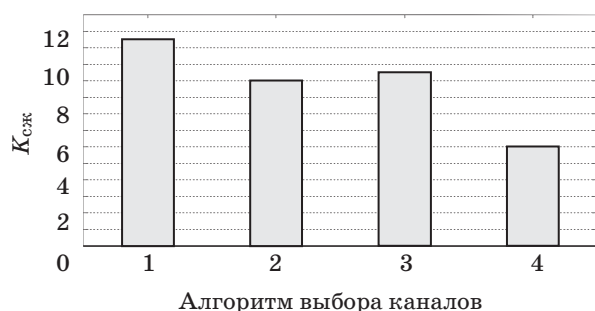
Будем называть наиболее информативными спектральные каналы, которые при корреляционном распознавании объектов по спектральным образам позволяют обеспечить вероятность распознавания, близкую к вероятности распознавания по полному объему многоспектральных данных. С учетом спектральной размерности многоспектральных данных оптических наблюдений выбор наиболее информативных спектральных каналов эквивалентен сжатию исходных видеоданных в десятки раз, хотя, строго говоря, это является сжатием с потерями — информация, содержащаяся в отброшенных спектральных каналах, теряется, по крайней мере, для обработки, которая осуществляется только с использованием видеоданных переданных спектральных каналов. Кроме того, наибольшие возможности

сжатия гиперспектральных данных за счет выбора для передачи наиболее информативных спектральных каналов открываются при наличии априорной информации о наблюдаемых объектах и используемом алгоритме их распознавания. Для различных объектов и алгоритмов распознавания наиболее информативные спектральные каналы будут различными.

В общем случае выбор наиболее информативных спектральных каналов осуществляется путем оценивания возможного вклада видеоданных каждого спектрального канала в решение конкретной задачи ДЗЗ и поиска такой комбинации спектральных каналов, которые в совокупности дают наилучший результат в смысле выбранного критерия. Простейшим методом сокращения числа передаваемых спектральных каналов является их равномерное прореживание, практически не требующее априорной информации о наблюдаемых объектах. Если же известны спектральные образы объектов, предполагаемых к распознаванию по спектральным признакам, то могут быть реализованы оптимальные и квазиоптимальные алгоритмы выбора наиболее информативных спектральных каналов.

Математическая постановка задачи выбора наиболее информативных спектральных каналов при корреляционном распознавании объектов по спектральным признакам формулируется как задача обнаружения координат существенного изменения свойств зарегистрированных спектральных характеристик. Участки существенного изменения свойств (скачков) анализируемых спектральных характеристик ищутся с использованием методов статистического оценивания в виде решающей статистики для обнаружения и оценки максимального правдоподобия координаты (спектрального канала) изменения свойств (скачков) анализируемого процесса [17, 19]. Известны также квазиоптимальные и эвристические алгоритмы выбора наиболее информативных спектральных каналов: с использованием вейвлет-анализа, по приращениям и по максимуму контраста спектральных образов [20].

Представлены оценки коэффициента сжатия $K_{сж}$ для четырех различных вариантов выбора наиболее информативных спектральных каналов гиперспектральных данных (рис. 4). В качестве тестового изображения использовался тот же тестовый гиперспектральный снимок, для которого на рис. 3 приведены коэффициенты сжатия при использовании программных алгоритмов сжатия видеоданных. Коэффициент сжатия определялся из условия снижения не более чем на 5 % вероятности корреляционного распознавания наблюдаемого объекта по ограниченному числу спектральных каналов по сравнению с распознаванием по полному объему гиперспектраль-



■ **Рис. 4.** Оценки коэффициента сжатия $K_{сж}$ для алгоритмов выбора спектральных каналов: 1 — оптимальный алгоритм; 2 — алгоритм анализа приращений; 3 — алгоритм на основе вейвлет-анализа; 4 — алгоритм равномерного прореживания

ных данных. Оптимальный алгоритм 1 обеспечивает уменьшение объема передаваемых видеоданных до 12,5 раза, квазиоптимальные алгоритмы 2 и 3 — уменьшение до 10 раз, простейший алгоритм 4 — до шести раз.

Уменьшение требуемой скорости передачи информации при использовании алгоритмов сжатия видеоданных определяется результирующим коэффициентом сжатия. При совместном использовании программных алгоритмов сжатия видеоданных в каждом спектральном канале и выбора наиболее информативных спектральных каналов из всего «гиперкуба» зарегистрированных гиперспектральных видеоданных возможно их сжатие в несколько десятков раз. Так, применение оптимального алгоритма выбора наиболее информативных спектральных каналов к тестовому гиперспектральному снимку объемом 4,8 Гбайт с числом спектральных каналов $b = 500$, для которого определялись коэффициенты сжатия, приведенные на рис. 3 и 4, позволило сократить число спектральных каналов до $b = 40$. Последующее применение к видеоданным этих спектральных каналов процедуры сжатия с использованием алгоритмов сжатия без потерь, таких как WinRar или WinZip, позволило дополнительно сжать данные еще в 1,5 раза. Результирующий коэффициент сжатия составляет 18,75, при этом практически отсутствуют потери в качестве решения задач тематической обработки видеоданных, а для их передачи при том же времени передачи, что при передаче со скоростью 100–150 Мбит/с всего объема гиперспектральных данных, может быть использован радиоканал со скоростью передачи информации 5–15 Мбит/с.

Заключение

В статье рассмотрены вопросы, связанные с согласованием информационной производительности бортовых гиперспектральных ОЭС

космических систем ДЗЗ с пропускной способностью спутниковых радиоканалов передачи данных на наземные ППИ. Полученные оценки информационной производительности существующих гиперспектрометров показывают, что суммарные объемы регистрируемых гиперспектральных видеоданных уже сейчас достигают доли терабайт, и в ближайшей перспективе неизбежно их увеличение. Для передачи с КА ДЗЗ на наземные ППИ полного потока зарегистрированных гиперспектральных данных необходимо достижение в спутниковых радиотехнических системах передачи информации скоростей передачи в единицы гигабит в секунду и выше, что превосходит возможности как существующих, так и перспективных высокоскоростных систем передачи информации, включая лазерные. Использование же параллельной многоканальной передачи означает, по сути, одновременную загрузку нескольких радиолиний и не всегда может быть технически реализовано. Отсюда следует вывод о необходимости сжатия регистрируемых гиперспектральных видеоданных для согласования информационной производительности гиперспектральных ОЭС КА ДЗЗ с характеристиками имеющих ограниченную пропускную способность радиоканалов передачи информации на наземные ППИ.

В статье показано, что перспективным способом выхода из сложившейся ситуации является совместное использование программных алгоритмов сжатия видеоданных в каждом спектральном канале и выбора для оперативной передачи из всего «гиперкуба» зарегистрированных гиперспектральных видеоданных ограниченного числа наиболее информативных спектральных каналов. При этом возможно сжатие исходных гиперспектральных данных в несколько десятков раз с соответствующим уменьшением требуемой скорости передачи информации в радиоканале. В результате при предельной скорости передачи в существующих и перспективных спутниковых высокоскоростных системах передачи информации 300–600 Мбит/с возможно использование среднескоростных систем передачи информации со скоростью передачи 10–30 Мбит/с. При этом выбор спектральных каналов для оперативной передачи может осуществляться на этапе планирования съемки на основании априорной информации об объектах, предполагаемых к распознаванию по спектральным признакам. Ограничениями для уменьшения числа передаваемых каналов является спектральная размерность гиперспектральных данных, допустимые потери в качестве решения задач ДЗЗ при их тематической обработке и необходимость использования априорной информации об объектах наблюдения.

Литература

1. Савиных В. П., Цветков В. Я. Геоинформационный анализ данных дистанционного зондирования. — М.: Картгеоцентр — Геодезиздат, 2001. — 228 с.
2. Бакланов А. И. Системы наблюдения и мониторинга. — М.: БИНОМ. Лаборатория знаний, 2009. — 234 с.
3. Сорокин Ю. С. Космическая съемка Земли 2010–2011. Ежегодник. — М.: Радиотехника, 2010. — 490 с.
4. Schowengerdt R. A. Remote Sensing. Models and Methods for Image Processing. — N. Y.: Academic Press, 2007. — 558 p.
5. Рождествен В. Н. Оптико-электронные системы экологического мониторинга природной среды. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2002. — 528 с.
6. Мальцев Г. Н., Козинев И. А., Данилкин А. П. Космические системы и технологии многоспектрального дистанционного зондирования Земли // Информация и космос. 2010. № 1. С. 148–158.
7. Урличич Ю. М., Селин В. А., Емельянов К. С. О приоритетах практической реализации развития космической системы дистанционного зондирования Земли // Наука и технологии в промышленности. 2012. № 1. Ч. 2. С. 71–80.
8. Бондур В. Г. Современные подходы к обработке больших потоков гиперспектральной и многоспектральной аэрокосмической информации // Исследование Земли из космоса. 2014. № 1. С. 4–16. doi: 10.7868/S0205961414010035
9. Козинев И. А., Бабанин И. Г., Богомазов А. Ю. Оперативная передача видеоданных с космических аппаратов дистанционного зондирования Земли // Известия Юго-Западного гос. ун-та. Сер. Управление, вычислительная техника, информатика. 2012. № 2–3. С. 102–105.
10. Гершензон В. Е., Кучейко А. А. Стандартизация оборудования станций приема данных дистанционного зондирования Земли // Пространственные данные. 2006. № 1. С. 33–43.
11. Шевчук Р. Б. Комплексы приема информации с российских спутников дистанционного зондирования Земли // Геоматика. 2012. № 2. С. 66–76.
12. Мальцев Г. Н. Современное состояние и перспективы развития многофункциональной космической системы ретрансляции // Изв. ГУАП. Аэрокосмическое приборостроение. Вып. 4. 2013. С. 13–18.
13. Носенко Ю. И., Лошкарев П. А. Единая территориально-распределенная информационная система дистанционного зондирования Земли — проблемы, решения, перспективы. Ч. 1 // Геоматика. 2010. № 3. С. 35–43.
14. Miano J. Compressed Image File Formats: JPEG, PNG, GIF, XBM, BMP. — Boston: (ACM Press) Addison-Wesley Professional, 1999. — 288 p.
15. Motta G., Rizzo F., Storer J. A. Hyperspectral Data Compression. — N. Y.: Springer Science and Business Media, 2006. — 421 p.
16. Фатеев В. Ф., Миньков С. А. Потенциальные возможности космических многоспектральных оптико-электронных приборов при обнаружении малоразмерных объектов // Оптический журнал. 2000. № 7. С. 5–11.
17. Мальцев Г. Н., Козинев И. А., Фатеев В. Ф. Методы выбора наиболее информативных спектральных каналов при дистанционном зондировании Земли с малых КА // Изв. вузов. Приборостроение. 2007. № 6. С. 23–31.
18. Тарасов В. В., Якушенков Ю. Г. Многоспектральные оптико-электронные системы // Специальная техника. 2002. № 4. С. 56–62.
19. Козинев И. А., Мальцев Г. Н. Модифицированный алгоритм обнаружения разладки случайного процесса и его применение при обработке многоспектральных данных // Информационно-управляющие системы. 2012. № 3. С. 9–17.
20. Козинев И. А. Обнаружение локальных свойств анализируемых сигналов и процессов с использованием вейвлет-преобразования // Информационно-управляющие системы. 2015. № 1. С. 21–28. doi:10.15217/issn1684-8853.2015.1.21

UDC 528.851

doi:10.15217/issn1684-8853.2016.2.74

Transfer of Hyperspectral Video Data of Earth Remote Sensing via Radio Channels of Limited Capacity

Maltsev G. N.^a, Dr. Sc., Tech., Professor, georgy_maltsev@mail.ruKozinov I. A.^a, PhD, Tech., Associate Professor, garry-spb@yandex.ru^aA. F. Mozhaiskiy Military Space Academy, 13, Zhdanovskaia Emb., 197198, Saint-Petersburg, Russian Federation

Introduction: Simultaneous use of hyperspectral imagery in hundreds of spectral channels is a steady tendency in the development of space systems for Earth remote sensing. It leads to anticipating growth of the informativeness of onboard optical-electronic observation systems in relation to the capacity of radio channels of information transfer from the spacecrafts. The problem is that the modern radio engineering information transmission systems with limited throughput cannot transfer complete volumes of registered hyperspectral video data either by direct operational transmission or by reproduction from an onboard storage device. It is necessary to develop technologies for the abbreviation (compression) of a hyperspectral video data flow for its transmission to ground stations of information reception, using satellite radio engineering information transmission systems with limited throughput. **Results:** The analysis of information productivity of the existing onboard hyperspectral optical-electronic systems showed that it reaches gigabits per second and will increase in the long term. At the same time, the throughput of radio channels of video data transmission from spacecrafts, taking into account the opportunities of the used frequency ranges and technical restrictions, does not exceed several

hundred megabits per second. To reduce the transmitted data amount, we suggest to share the algorithms of video data shrinking in each spectral channel and choose only several most informative spectral channels out of all the registered hyperspectral video data. This will reduce the transmitted data amount in several tens times and coordinate the information productivity of optical-electronic observation systems with the throughput of the radio channels. Achievable compression ratios are defined for hyperspectral video data when sharing different compression methods and methods of choosing the most informative spectral channels. **Practical relevance:** The obtained results can be applied in the design and use of space systems of Earth remote sensing with hyperspectral shooting for choosing technologies and modes to transmit the spacecraft observation data via radio channels of limited capacity.

Keywords — Remote Sensing of Earth, Hyper Spectrometer, Spectral Channel, Information Transfer Speed, Redundancy Reduction, Video Data Compression.

References

1. Savinykh V. P., Tsvetkov V. Y. *Geoinformatsionyi analiz dannykh distantsionnogo zondirovaniia* [Geoinformational Analysis of Remote Sensing Data]. Moscow, Kartgeotsentr — Geodezizdat Publ., 2001. 228 p. (In Russian).
2. Baklanov A. I. *Sistemy nabludenii i monitoringa* [System Surveillance and Monitoring]. Moscow, BINOM. Laboratoriia znaniia Publ., 2009. 234 p. (In Russian).
3. Sorokin U. S. *Kosmicheskaiia s"emka Zemli 2010–2011. Ezhegodnik* [Earth Remote Sensing. Yearbook]. Moscow, Radio-tekhnika Publ., 2010. 490 p. (In Russian).
4. Schowengerdt R. A. *Remote Sensing. Models and Methods for Image Processing*. New York, Academic Press, 2007. 558 p.
5. Rozhdestvin V. N. *Optikoelektronnye sistemy ekologicheskogo monitoringa prirodnoi sredy* [Optical-Electronic Systems of Environmental Monitoring of Environment]. Moscow, MGTU im. N. E. Bauman Publ., 2002. 528 p. (In Russian).
6. Maltsev G. N., Kozinov I. A., Danilkin A. P. Space Systems and Technologies Interned for Multispectral Remote Sensing of the Earth. *Informatsiia i kosmos*, 2010, no. 1, pp. 148–158 (In Russian).
7. Urlichich Y. M., Selin V. A., Emel'yanov S. K. On the Priorities of Practical Implementation of Development of Space System of Remote Sounding of the Earth. *Nauka i tekhnologii v promyshlennosti*, 2012, no. 1–2, pp. 71–80 (In Russian).
8. Bondur V. G. Modern Approaches to Processing of Big Flows of Hyper Spectral and Multispectral Space Information. *Issledovanie Zemli iz kosmosa*, 2014, no. 1, pp. 4–16 (In Russian) doi: 10.7868/S0205961414010035
9. Kozinov I. A., Babanin I. G., Bogomazov A. Y. Rapid Transmission of Video Data from the Spacecraft for Remote Sensing of the Earth. *Izvestiia Iugo-Zapadnogo gosudarstvennogo universiteta. Ser. Upravlenie, vychislitel'naia tekhnika, informatika*, 2012, no. 2–3, pp. 102–106 (In Russian).
10. Gershenzon V. E., Kucheiko A. A. Standardization of the Equipment of Stations of Reception of Data of Remote Sensing of Earth. *Prostranstvennye dannye*, 2006, no. 1, pp. 33–43 (In Russian).
11. Shevchuk R. Complexes for Data Receiving from Russian Remote Sensing Satellites. *Geomatika*, 2012, no. 2, pp. 66–76 (In Russian).
12. Maltsev G. N. Current State and Prospects of Development of Multifunctional Space System Relay. *Izvestiia GUAP. Aerokosmicheskoe priborostroenie*, 2013, iss. 4, pp. 13–18 (In Russian).
13. Nosenko Y. I., Loshkarev P. A. Unified Geographically Distributed Information System of Earth Remote Sensing — Problems, Solutions, Perspectives. Part 1. *Geomatika*, 2010, no. 3, pp. 35–43 (In Russian).
14. Miano J. *Compressed Image File Formats: JPEG, PNG, GIF, XBM, BMP*. Boston, (ACMPress) Addison-Wesley Professional, 1999. 288 p.
15. Motta G., Rizzo F., Storer J. A. *Hyperspectral Data Compression*. New York, Springer Science and Business Media, 2006. 421 p.
16. Fateev V. F., Minkov S. A. The Potential of Space Multi-Spectral Opto-Electronic Devices in the Detection of Small Objects. *Opticheskii zhurnal*, 2000, no. 7, pp. 5–11 (In Russian).
17. Maltsev G. N., Kozinov I. A., Fateev V. F. Methods of Choice of the Most Informational Spectral Channels under Remote Flexing of the Earth from the Small Spacecrafts. *Izvestiia vuzov. Priborostroenie*, 2007, no. 6, pp. 23–31 (In Russian).
18. Tarasov V. V., Yakushenkov Yu. G. Multi-Spectral Optical-Electronic Systems. *Spetsial'naia tekhnika*, 2002, no. 4, pp. 56–62 (In Russian).
19. Kozinov I. A., Maltsev G. N. Modified Algorithm of the Detection of Abrupt Changes in a Casual Process and its use for Processing of Multispectral Data. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2012, no. 3, pp. 9–17 (In Russian).
20. Kozinov I. A. Detecting Local Characteristic of Analyzed Signals and Processes using Wavelet Transformation. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 1, pp. 21–28 (In Russian). doi:10.15217/issn1684-8853.2015.1.21

МЕТОДЫ АКУСТООПТИКИ В ИССЛЕДОВАНИИ СВОЙСТВ МАТЕРИАЛОВ ДЛЯ АКУСТООПТИЧЕСКИХ И АКУСТОЭЛЕКТРОННЫХ УСТРОЙСТВ

С. В. Кулаков^а, доктор техн. наук, профессор

О. Л. Балышева^а, канд. техн. наук, доцент

В. В. Клузгин^а, доктор техн. наук, профессор

^аСанкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ

Постановка проблемы: монокристаллы для акустооптических и акустоэлектронных устройств должны обладать свойствами, обеспечивающими практическое применение распространяющихся в них оптических и акустических волн для обработки сигналов. Для исследования свойств и измерения различных параметров монокристаллов используются методы, многие из которых требуют применения специального оборудования и дополнительных средств отображения результатов. **Цель:** показать эффективность и наглядность использования акустооптического метода для исследования структуры акустических полей, акустического затухания и измерения нелинейных параметров кристаллов. **Методы:** оригинальность подхода состоит в использовании акустооптического взаимодействия как для комплексного исследования свойств монокристаллов, так и для визуализации с помощью «теневых картин» (Schlierens images). **Результаты:** оценена дифракционная расходимость и измерено отклонение направления переноса энергии от волновой нормали. Одновременная визуализация продольной и сдвиговой акустических мод позволила сравнить их дифракционную расходимость и измерить углы отклонения направления переноса энергии от волновой нормали для каждой из мод. Получена картина пространственного распределения мощности акустической волны, позволяющая выявить ростовую неоднородность кристалла. Выполнены измерения параметров упругой нелинейности. Показано, что уровень затухания в упруго-нелинейной среде зависит от амплитуды волны. **Практическая значимость:** акустооптический подход позволяет выполнять визуальную оценку качества исследуемых кристаллов и их пригодности для построения устройств, а также оценивать затухание акустических волн и степень проявления упругой нелинейности.

Ключевые слова — акустооптика, акустическая волна, кристалл, «теневые картины», затухание, упругая нелинейность.

Введение

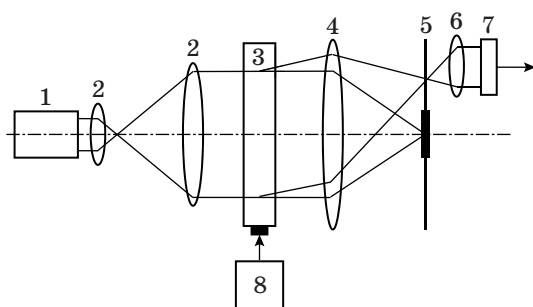
Акустоэлектронные и акустооптические устройства широко используются в аппаратуре современных информационно-измерительных систем. Одним из перспективных и востребованных применений таких устройств служат датчики в составе различных систем мониторинга — от технического состояния промышленных объектов до состояния здоровья человека. Особенностью устройств акустоэлектроники и акустооптики является необходимость применения специальных материалов — монокристаллов. Используемые монокристаллы должны удовлетворять ряду требований [1, 2], поэтому необходимо детальное исследование их свойств. Интересно, что инструментом исследований монокристаллов, применяемых в акустооптических и акустоэлектронных устройствах, может служить также акустооптическое устройство.

В дополнение к широким возможностям по обработке информации акустооптические методы являются эффективным средством различных физических измерений [3, 4], в частности средством исследования свойств материалов. С помощью акустооптики возможны измерения параметров и качественные исследования акусти-

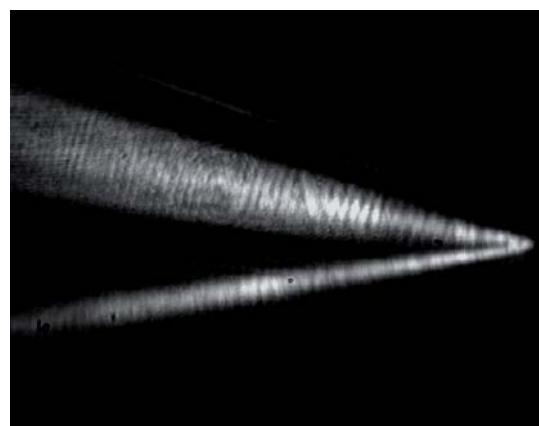
ческих полей (скорости, затухания, структуры акустических мод), оптических полей (например, оптической однородности образцов многокомпонентных кристаллов [5]) и характеристик самой среды взаимодействия — кристаллов. Для среды возможно исследовать анизотропные свойства, дифракционные и нелинейные эффекты, выявлять коллимирующие направления. В данной статье продемонстрировано применение акустооптического метода для исследования структуры акустических полей, акустического затухания и измерения нелинейных параметров кристаллов, используемых в акустооптических и акустоэлектронных устройствах.

Оптическая визуализация структуры акустических полей

Акустооптическое взаимодействие можно использовать для визуализации особенностей распространения и структуры акустических волн в кристаллах. На рис. 1 представлена схема визуализации. В системе с помощью диафрагмы 5 выделяется (фильтруется) оптический дифракционный порядок, затем изображение этого порядка восстанавливается. ССД-камера и компьютерный дисплей регистрируют изображение. Метод



■ **Рис. 1.** Схема визуализации акустических полей: 1 — источник оптического излучения ($\lambda = 0,63$ мкм); 2 — коллимирующая система; 3 — акустооптическая ячейка; 4 — оптическая линза; 5 — диафрагма; 6 — линза изображения; 7 — ССД-камера; 8 — генератор акустического сигнала



■ **Рис. 3.** Визуализация продольной и сдвиговой волн одновременно в кристалле GaP

«теневых картин» (Schlieren image) является очень чувствительным, позволяя накапливать оптический сигнал и осуществлять последетекторную обработку. Этот метод можно использовать для наглядной и численной оценки акустического затухания, эффектов дифракционной расходимости (меры энергетической расходимости), условий отражения, преломления и трансформации акустических мод.

Структура акустической сдвиговой волны, распространяющейся в плоскости (001) кристалла двуокиси теллура (TeO_2), показана на рис. 2, причем волновая нормаль здесь была смещена на $1,18^\circ$ относительно направления [110] (вертикальное направление на рис. 5), а угол отклонения направления переноса энергии от волновой нормали достигал $\sim 44^\circ$.

В представленной картине очевидны проявления дифракционной расходимости и затухания этой моды.

Уникальные возможности акустооптического метода визуализации структуры акустических полей особенно важны для анизотропных мате-



■ **Рис. 2.** Структура сдвиговой волны в плоскости (001) кристалла TeO_2

риалов. В качестве иллюстрации на рис. 3 «теневая картина» изображает структуру и особенности распространения двух акустических волн в кристалле фосфида галлия (GaP). И продольная, и сдвиговая волны излучаются одним пьезопреобразователем с волновой нормалью, ориентированной вдоль горизонтальной оси рисунка. Для обеих акустических мод наблюдается отклонение направления переноса энергии от волновой нормали с разными знаками. Кроме того, выделяются различия в энергетической расходимости этих волн, что связано с разной мерой и качеством упругой анизотропии для разных акустических мод.

Измерения затухания акустических волн

Для идеальных пространственно однородных сред коэффициент затухания α определяет пространственную зависимость амплитуды акустической волны в форме

$$s(x) = s_0 \exp(-\alpha x),$$

где s_0 — начальное значение амплитуды, а x — координата, вдоль которой распространяется акустическая волна. Реальные материалы (в том числе упорядоченные среды, кристаллы) обладают пространственными неоднородностями, возникновение которых определяется различными технологическими факторами. Эти неоднородности нарушают периодичность микростроения среды и могут быть вызваны различными примесями, включениями, дислокациями и т. д. Если задать функцию распределения неоднородностей вдоль координаты x (для упрощения рассмотрим одномерный вариант) $N(x)$, то коэффициент затухания $\alpha(x)$ будет связан с этим распределением в простейшей форме

$$\alpha(x) = \alpha_0 + \beta N(x),$$

а пространственное распределение амплитуды акустической волны в логарифмическом масштабе имеет вид

$$\ln \frac{s(x)}{s_0} = -\alpha(x)dx.$$

В акустооптических измерениях уровень оптического сигнала пропорционален мощности акустической волны, поэтому удобнее пользоваться модифицированным коэффициентом затухания $\alpha^*(x)$, определенным в форме

$$\alpha^*(x) = 20 \lg_e \alpha(x),$$

где e — основание натурального логарифма.

В общем случае для реальных сред распределение неоднородностей $N(x)$ является случайной функцией, причем вероятностный закон распределения величины N вследствие центральной предельной теоремы можно считать нормальным:

$$p(N) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(N - N^*)^2}{2\sigma^2} \right],$$

где N^* — среднее значение концентрации неоднородностей среды, а σ — дисперсия концентрации, величина которой связана со средним значением, причем степень связи определяется условиями синтеза материала. В этом случае коэффициент затухания состоит из двух компонент:

$$\alpha(x) = \alpha_0 + \gamma(x),$$

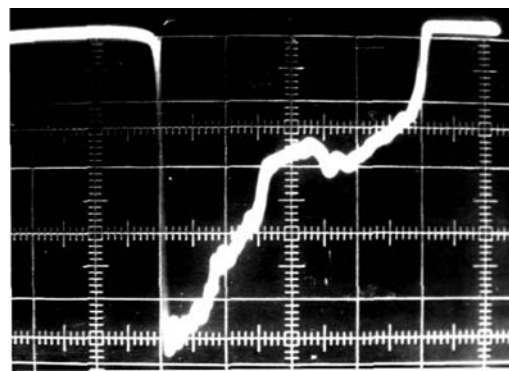
где α_0 — коэффициент затухания идеальной среды, не содержащей неоднородностей, а $\gamma(x)$ — введенная неоднородностями часть затухания. Вероятностный закон распределения логарифмической меры нормированных значений амплитуды акустической волны определяется аналогичным образом, причем модульные значения моментов этого распределения увеличиваются с ростом координаты x . Истинные логарифмические значения моментов являются отрицательными величинами.

В неоднородных средах потенциальная минимальная погрешность измерения затухания ограничена соотношением

$$\Delta = \frac{\beta\sigma}{\beta N^* + \alpha_0}$$

и не может быть уменьшена какими-либо инструментальными методами. Поэтому в реальных средах минимальный уровень затухания, который может быть измерен, всегда ограничен.

Осциллограмма сигнала фотоприемника (рис. 4) иллюстрирует особенности измерения затухания в неоднородных средах. Здесь использовался кристалл GaP с явно выраженной ростовой неоднородностью в центральной части.



■ Рис. 4. Пространственное распределение мощности акустической волны в неоднородном образце

Практически наибольшее влияние на затухание оказывают структурные неоднородности реальных сред и нелинейные эффекты, присущие всем материалам. Наличие микрон неоднородностей материала может привести к заметной деформации пространственных распределений амплитуды акустической волны и существенно усложнить процесс определения коэффициента затухания. Случайный характер распределения неоднородностей в объеме среды ограничивает потенциальную точность определения затухания, а также минимальный уровень коэффициента затухания, который может быть измерен.

Значительное влияние на характеристики затухания оказывает упругая нелинейность материала. В общем случае нелинейность не только увеличивает коэффициент затухания в зависимости от условий измерения, но и меняет форму пространственных распределений амплитуды волны. Степень влияния упругой нелинейности на затухание зависит от уровня мощности акустической волны, поэтому для минимизации влияния нелинейности необходимо проводить измерения при минимально возможной амплитуде.

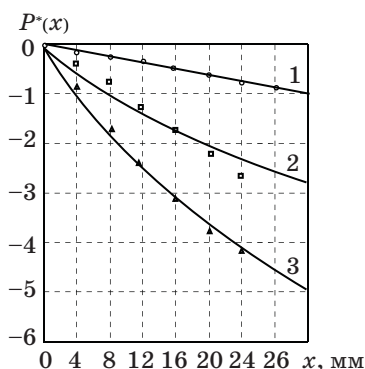
Измерение параметров упругой нелинейности материалов

Явление упругой нелинейности является одним из важнейших физических факторов, ограничивающим и искажающим характеристики и параметры акустооптических устройств. Вследствие нелинейности среды увеличивается коэффициент затухания акустической волны, причем в некоторых материалах это увеличение по экспериментальным данным может быть значительным (в 5–7 раз, например, в кристаллах LiNbO_3 и $\text{NaBi}(\text{MoO}_4)_2$). Характерно, что величина затухания в реальных кристаллах зависит от уровня плотности акустической мощности, скорости изменения сигнала во времени (частоты сигнала) и нелинейных характеристик материала.

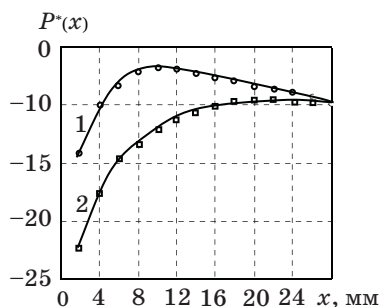
На рис. 5 представлены логарифмические пространственные распределения амплитуды (мощности) основной моды (продольной акустической волны), распространяющейся вдоль направления [001] в кристалле $\text{NaBi}(\text{MoO}_4)_2$ при разных уровнях плотности мощности на грани возбуждения. Здесь наглядно представлено влияние упругой нелинейности, проявление которой тем больше, чем выше плотность мощности акустического сигнала. Если при малой мощности коэффициент затухания составляет 0,32 дБ/см при частоте гармоники $f = 270$ МГц, то при большой мощности коэффициент затухания возрастает почти на порядок.

На рис. 6 представлены совместно с экспериментальными данными расчетные зависимости, иллюстрирующие логарифмические пространственные распределения мощности второй гармоники в кристалле $\text{NaBi}(\text{MoO}_4)_2$ при разных значениях начальной плотности мощности основной моды на границе возбуждения волны.

Акустооптическим методом экспериментально исследовались характеристики упругой нели-



■ Рис. 5. Пространственные распределения мощности основной гармоники $P^*(x) = P_1(x)/P_1(0)$ при разных уровнях мощности $P_1(0)$ на границе возбуждения: $P_1(0) = 10$ мВт/см² (кривая 1); $P_1(0) = 3$ Вт/см² (кривая 2); $P_1(0) = 6$ Вт/см² (кривая 3)



■ Рис. 6. Пространственные распределения мощности второй гармоники $P^*(x) = P_2(x)/P_1(0)$ при разных значениях мощности возбуждения: $P_1(0) = 6$ Вт/см² (кривая 1); $P_1(0) = 3$ Вт/см² (кривая 2)

нейности кристаллов LiNbO_3 , TeO_2 , KRS-5, $\text{NaBi}(\text{MoO}_4)_2$, $\text{Bi}_{12}\text{GeO}_{20}$, PbMoO_4 и других материалов. В средах с небольшим коэффициентом затухания в номинальных условиях и при последовательной подстройке угла Брэгга оказалось возможным наблюдать ряд высших гармоник вплоть до 15–20 включительно, что в определенных условиях соответствовало частотам гармоник 7–12 ГГц.

Уровень затухания акустической волны в упруго-нелинейной среде зависит от амплитуды волны и в этой зависимости может меняться в несколько раз. Поэтому при оценке естественного затухания необходимо производить измерения при минимально возможной мощности возбуждаемой волны.

В общем случае пространственные распределения волны, измеряемые экспериментально, отклоняются от экспоненциальной зависимости вследствие влияния нелинейных эффектов. Поэтому оценка коэффициента затухания оказывается приблизительной.

Заключение

Показано, что акустооптический метод служит эффективным и наглядным средством как качественных исследований, так и количественных измерений характеристик кристаллов. Наблюдаемые структурные изменения и искажения акустических мод в кристаллах могут иметь значительное влияние на характеристики реализуемых акустооптических и акустоэлектронных устройств. Метод позволяет определять частотный характер затухания при использовании возбуждения пьезопреобразователя на высших гармониках. С помощью акустооптического взаимодействия выявляются характеристики и параметры упругой нелинейности среды взаимодействия — монокристаллов. Наличие упругой нелинейности приводит к искажению гармонической формы акустической волны и, соответственно, к генерации высших гармоник.

Литература

1. Балышева О. Л. и др. Акустоэлектронные устройства обработки и генерации сигналов. Принципы работы, расчета и проектирования / О. Л. Балышева, В. И. Григорьевский, Ю. В. Гуляев, В. Ф. Дмитриев, Г. Д. Мансфельд: монография; под ред. академика РАН Ю. В. Гуляева. — М.: Радиотехника, 2012. — 576 с.
2. Балышева О. Л. Критерии выбора материалов для датчиков на поверхностных акустических волнах // Датчики и системы. 2015. № 2. С. 23–27.

- | | |
|--|---|
| <p>3. Гусев О. Б., Клудзин В. В. Акустооптические измерения. — Л.: Изд-во ЛГУ, 1987. — 152 с.</p> <p>4. Sergei V. Kulakov, et al. Acousto-Optics as an Efficient Method for Physical Measurements / Sergei V. Kulakov, Olga L. Balysheva, Arcenii Yu. Zhdanov, Victor V. Kludzin, Oleg V. Shakin // <i>Physics Procedia</i>. 2015. N 70. P. 253–275.</p> | <p>5. Бальшева О. Л., Клудзин В. В., Кулаков С. В., Дмитриев В. Ф. Материалы группы лангасита для акустоэлектронной элементной базы современных информационно-коммуникационных систем // <i>Информационно-управляющие системы</i>. 2012. № 6. С. 67–72.</p> |
|--|---|

UDC 534:621.382

doi:10.15217/issn1684-8853.2016.2.84

Methods of Acousto-Optics in Studying the Properties of Materials for Acousto-Optic and Acousto-Electronic Devices

Kulakov S. V.^a, Dr. Sc., Tech., Professor, svk25@mail.ruBalysheva O. L.^a, PhD, Tech., Associate Professor, balysheva@mail.ruKludzin V. V.^a, Dr. Sc., Tech., Professor^aSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

Introduction: Monocrystals for acousto-optic and acousto-electronic devices should have properties which provide practical application of their acoustic and optical waves for signal processing. To study the properties and measurement of various parameters of monocrystals, various methods are used, many of which require special equipment and additional effort to display the results. **Purpose:** The goal of this paper is to show the efficiency and easy visualization of using the acousto-optic method for studying the structure of acoustic fields, acoustic attenuation and measurement of nonlinear parameters of crystals. **Methods:** The originality of the approach is using acousto-optic interaction both for a complex study of the properties of monocrystals and for the visualization through “Schlierens images”. **Results:** The degree of diffraction was evaluated, and the power flow angle was measured. Simultaneous visualization of longitudinal and shear acoustic modes provided the opportunity to compare their diffraction divergence and to measure the power flow angle for each mode. A picture was obtained for the spatial distribution of the power of acoustic waves in crystals, and the parameters of the elastic nonlinearity were measured. It was shown that the level of elastic damping in nonlinear medium depends on the wave amplitude. **Practical relevance:** The acousto-optic method provides a visual assessment of the quality of the crystals and their suitability for constructing devices. Moreover, the acousto-optic approach allows you to estimate the attenuation of acoustic waves and the intensity of the elastic nonlinearity.

Keywords — Acousto-Optics, Acoustic Wave, Crystal, Schlierens Images, Attenuation, Nonlinearity Parameters.

References

- | | |
|--|---|
| <p>1. Balysheva O. L., Grigorievsky V. I., Gulyaev Yu. V., Dmitriev V. F., Mansfeld G. D. <i>Akustoelektronnye ustroystva obrabotki i generatsyi signalov. Printsipy raboty, rasceta i proektirovaniia</i> [Acoustoelectronic Devices for Signal Processing and Generation. Principles of Operation Modeling and Development]. Moscow, Radiotekhnika Publ., 2012. 576 p. (In Russian).</p> <p>2. Balysheva O. L. Criteria of a Choice of Materials for Sensors on Surface Acoustic Waves. <i>Datchiki i sistemy</i>, 2015, no. 2(189), pp. 23–27 (In Russian).</p> <p>3. Gusev O. B., Kludzin V. V. <i>Akusroopticheskie izmereniia</i> [Acousto-Optic Measurements]. Leningrad, LGU Publ., 1987. 152 p. (In Russian).</p> | <p>4. Sergei V. Kulakov, Olga L. Balysheva, Arcenii Yu. Zhdanov, Victor V. Kludzin, Oleg V. Shakin. Acousto-Optics as an Efficient Method for Physical Measurements. <i>Physics Procedia</i>, 2015, no. 70, pp. 253–275.</p> <p>5. Balysheva O. L., Kludzin V. V., Kulakov S. V., Dmitriev V. F. Langasite Group Materials for Acoustic-Electronic Elemental Base in Modern Information and Communication Systems. <i>Informatsionno-upravliaiushchie sistemy</i> [Information and Control Systems], 2012, no. 6(61), pp. 67–72 (In Russian).</p> |
|--|---|

АНАЛИЗ АЛГОРИТМА КОМПЛЕКСНОЙ ОБРАБОТКИ УГЛОМЕРНОЙ ИНФОРМАЦИИ В НАВИГАЦИОННОЙ СИСТЕМЕ ПОДВИЖНОГО ОБЪЕКТА

А. Ю. Каплин^а, канд. техн. наук

М. Г. Степанов^а, доктор техн. наук, профессор

^аОАО «Радиоавионика», Санкт-Петербург, РФ

Постановка проблемы: при комплексировании инерциальных датчиков, входящих в состав системы навигации подвижного объекта, с навигационными датчиками других типов: цифровым магнитным компасом, модулем спутниковой навигации, другими приборами, — последние используются в качестве внешнего дополнительного источника информации для выделения и оценки по алгоритму калмановской фильтрации систематических (медленно меняющихся) ошибок инерциальных датчиков с их последующей коррекцией по разомкнутой или замкнутой схеме. В результате точность навигационной системы объекта определяется точностью оценок указанных ошибок и слабо зависит от информационного вклада измерений других датчиков. **Цель:** разработать и проанализировать алгоритм комплексной обработки угломерной информации в навигационной системе подвижного объекта. **Результаты:** применительно к угломерному каналу навигационной системы подвижного объекта предложен и проанализирован алгоритм комплексной обработки полного объема измерительной информации, построенный на основе фильтра Калмана. Рассмотрен и обоснован вклад отдельных компонентов алгоритма в общий процесс обработки. Продемонстрирована работа механизма весового суммирования, реализуемая калмановским алгоритмом фильтрации угломерных данных. Описаны варианты практического задания начальных условий для фильтра. **Практическая значимость:** сформулированный алгоритм комплексной обработки может использоваться как основа специального программного обеспечения навигационной системы подвижных наземных и околоземных объектов. С методической точки зрения полученные результаты полезны молодым специалистам для понимания и освоения прикладных аспектов калмановской фильтрации.

Ключевые слова — комплексная обработка, фильтр Калмана, угломерный канал, навигационная система, подвижный объект.

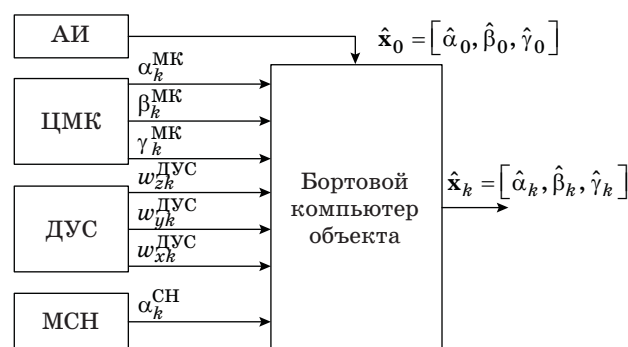
Введение

Статья опирается на принципы и алгоритмы, изложенные в основополагающих работах [1–4], а также прикладных статьях [5–8]. По замыслу и содержанию статья является обобщением результатов работы [9]. В доступной форме в ней применительно к конкретной прикладной задаче анализируется механизм комплексной весовой обработки данных различных угломерных датчиков, реализуемый алгоритмом калмановской фильтрации.

Объектом рассмотрения является угломерный канал навигационной системы подвижного наземного (околоземного) объекта, распространенный состав датчиков которого включает трехосевой цифровой магнитный компас (ЦМК), трехосевой датчик угловых скоростей (ДУС) и модуль спутниковой навигации (МСН) [5, 10]. Датчики обеспечивают измерение углов ориентации объекта — азимута $\alpha_k^{МК}$, тангажа $\beta_k^{МК}$ и крена $\gamma_k^{МК}$, соответствующих угловых скоростей $w_{zk}^{ДУС}$, $w_{yk}^{ДУС}$, $w_{xk}^{ДУС}$ и путевого угла $\alpha_k^{СН}$. В состав канала обычно также входит трехосевой акселерометрический инклинометр (АИ), определяющий значения начальных (статических) углов ориентации объекта α_0 , β_0 , γ_0 , в отсчете от которых проводятся дальнейшие угловые измерения в ходе его движения. Функциональная схема

комплексной обработки приведена на рисунке (обозначения $\hat{\alpha}_0$, $\hat{\alpha}_k$ раскрываются далее).

Предполагается, что перед комплексной обработкой текущие измерения проходят предварительную обработку, обеспечивающую ввод систематических поправок (в частности, на магнитное склонение), независимое по измеряемым параметрам сглаживание и сжатие результатов и их привязку к единой шкале времени с одинаковым шагом $T_k = t_k - t_{k-1}$. Считается также, что предварительно проводится высокоточная калибровка ДУС, ЦМК и АИ [10, 11], позволяющая при обработке компенсировать смещения и асимметрию коэффициентов передачи чувствительных осей датчиков.



■ Функциональная схема комплексной обработки

Обоснование алгоритма комплексной обработки

В соответствии с положениями теории оптимальной линейной фильтрации в процессе обработки измерений оценке подлежит текущий вектор состояния $\mathbf{x}_k$ линейной динамической системы, описываемой уравнением (динамической моделью) вида

$$\mathbf{x}_k = \Phi \mathbf{x}_{k-1} + \mathbf{D} \mathbf{w}_{k-1} + \mathbf{G} \mathbf{n}_{k-1},$$

$$\text{н. у. } \mathbf{x}_0 = \mathbf{x}(0), \quad k = 1, 2, \dots, \quad (1)$$

где $\mathbf{n}_{k-1}$ — стандартный векторный гауссов белый шум с нулевым средним и единичной интенсивностью, $\mathbf{n}_{k-1} \sim \mathbf{N}(0, \mathbf{I})$; $\mathbf{w}_{k-1}$ — известный вектор управления; Φ , $\mathbf{D}$ — переходные матрицы состояния; $\mathbf{G}$ — матрица интенсивности шума $\mathbf{G} \mathbf{n}_{k-1}$.

Измерения описываются уравнением

$$\mathbf{y}_k = \mathbf{H} \mathbf{x}_k + \mathbf{G}_y \mathbf{v}_k, \quad (2)$$

где $\mathbf{v}_k \sim \mathbf{N}(0, \mathbf{I})$ — стандартный векторный шум измерения; $\mathbf{G}_y$ — матрица интенсивности шума $\mathbf{G}_y \mathbf{v}_k$; $\mathbf{H}$ — матрица измерения.

В этих условиях оптимальной по минимуму среднего квадрата ошибки оценивания (минимуму ковариационной матрицы ошибок) является рекуррентная процедура (фильтр Калмана)

$$\hat{\mathbf{x}}_k = \Phi \hat{\mathbf{x}}_{k-1} + \mathbf{P}_k \mathbf{H}^T (\mathbf{G}_y \mathbf{G}_y^T)^{-1} (\mathbf{y}_k - \mathbf{H} \Phi \hat{\mathbf{x}}_{k-1}) + \mathbf{P}_k \mathbf{P}_{k/k-1}^{-1} \mathbf{D} \mathbf{w}_{k-1}, \quad \text{н. у. } \hat{\mathbf{x}}_0, \quad (3)$$

$$\mathbf{P}_{k/k-1} = \Phi \mathbf{P}_{k-1} \Phi^T + \mathbf{G} \mathbf{G}^T;$$

$$\mathbf{P}_k = \left[\mathbf{P}_{k/k-1}^{-1} + \mathbf{H}^T (\mathbf{G}_y \mathbf{G}_y^T)^{-1} \mathbf{H} \right]^{-1} =$$

$$= \mathbf{P}_{k/k-1} - \mathbf{P}_{k/k-1} \mathbf{H}^T (\mathbf{H} \mathbf{P}_{k/k-1} \mathbf{H}^T + \mathbf{G}_y \mathbf{G}_y^T)^{-1} \mathbf{H} \mathbf{P}_{k/k-1},$$

$$\text{н. у. } \mathbf{P}_0,$$

где $\mathbf{P}_k$ — ковариационная матрица ошибок оценивания; $\mathbf{P}_{k/k-1}$ — экстраполированное на шаг значение этой матрицы.

Применительно к рассматриваемой прикладной проблеме алгоритм комплексной обработки угломерной информации решает задачу формирования по данным ЦМК, ДУС и МСН текущей калмановской оценки углов ориентации подвижного объекта.

В основу комплексирования положены следующие соображения (рассмотрение начнем с непрерывного времени t с последующим возвратом к дискретному времени t_k).

Между текущими значениями углов и угловых скоростей существует очевидная аналитическая зависимость

$$\frac{d\alpha(t)}{dt} = w_z(t), \quad \alpha_0 = \alpha(t_0);$$

$$\frac{d\beta(t)}{dt} = w_y(t), \quad \beta_0 = \beta(t_0);$$

$$\frac{d\gamma(t)}{dt} = w_x(t), \quad \gamma_0 = \gamma(t_0). \quad (4)$$

При известных функциях времени $w_z(t)$, $w_y(t)$, $w_x(t)$ (известном управлении) эти соотношения могли бы быть исходными для составления исчерпывающей динамической модели изменения углов $\alpha(t)$, $\beta(t)$, $\gamma(t)$. Однако, поскольку на практике истинные значения угловых скоростей $w_z(t)$, $w_y(t)$, $w_x(t)$ не известны, в соответствии с фундаментальной теоремой разделения и принципом распределения информации они заменяются на измеренные ДУС значения $w_z^{\text{ДУС}}(t)$, $w_y^{\text{ДУС}}(t)$, $w_x^{\text{ДУС}}(t)$ с последующим корректным учетом присущих им ошибок измерений.

Измеренные значения угловых скоростей представим в виде сумм

$$w_z^{\text{ДУС}}(t) = w_z(t) + \varepsilon_{wz}(t);$$

$$w_y^{\text{ДУС}}(t) = w_y(t) + \varepsilon_{wy}(t); \quad (5)$$

$$w_x^{\text{ДУС}}(t) = w_x(t) + \varepsilon_{wx}(t)$$

неизвестных истинных значений w_z , w_y , w_x и ошибок измерений ε_{wz} , ε_{wy} , ε_{wx} . В свою очередь, ошибки измерений запишем в виде

$$\varepsilon_{wx}(t) = \sigma_{wx} n_{wx}(t), \quad \varepsilon_{wy}(t) = \sigma_{wy} n_{wy}(t),$$

$$\varepsilon_{wz}(t) = \sigma_{wz} n_{wz}(t), \quad (6)$$

где n_{wz} , n_{wy} , n_{wx} — стандартные независимые белые гауссовы шумы, $\mathbf{n}_w \sim \mathbf{N}(0, \mathbf{I})$; $\sigma_{wx} = \sigma_{wy} = \sigma_{wz} = \sigma_w$ — среднеквадратические ошибки измерения угловых скоростей (полагаются равными для всех трех каналов ДУС).

Соотношения (4) с учетом (5) и (6) можно переписать в виде

$$\frac{d\alpha(t)}{dt} = w_z^{\text{ДУС}}(t) - \sigma_w n_{wz}(t);$$

$$\frac{d\beta(t)}{dt} = w_y^{\text{ДУС}}(t) - \sigma_w n_{wy}(t); \quad (7)$$

$$\frac{d\gamma(t)}{dt} = w_x^{\text{ДУС}}(t) - \sigma_w n_{wx}(t).$$

Переходя в (7) к конечным разностям, получим

$$\alpha_k = \alpha_{k-1} + T w_{z,k-1}^{\text{ДУС}} - T \sigma_w n_{wz,k-1};$$

$$\beta_k = \beta_{k-1} + T w_{y,k-1}^{\text{ДУС}} - T \sigma_w n_{wy,k-1}; \quad (8)$$

$$\gamma_k = \gamma_{k-1} + T w_{x,k-1}^{\text{ДУС}} - T \sigma_w n_{wx,k-1},$$

где $T = t_k - t_{k-1}$.

Сравнивая (1) и (8), убеждаемся, что они совпадают при введении следующих обозначений:

$$\mathbf{x}_k = \begin{bmatrix} \alpha_k \\ \beta_k \\ \gamma_k \end{bmatrix}, \quad \mathbf{\Phi} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{w}_{k-1} = \begin{bmatrix} w_{z,k-1}^{\text{ДУС}} \\ w_{y,k-1}^{\text{ДУС}} \\ w_{x,k-1}^{\text{ДУС}} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} T & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix},$$

$$\mathbf{n}_{k-1} = \begin{bmatrix} n_{wz,k-1} \\ n_{wy,k-1} \\ n_{wx,k-1} \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} -T\sigma_w & 0 & 0 \\ 0 & -T\sigma_w & 0 \\ 0 & 0 & -T\sigma_w \end{bmatrix}.$$

Обозначая

$$\mathbf{y}_k = \begin{bmatrix} \alpha_k^{\text{МК}} \\ \beta_k^{\text{МК}} \\ \gamma_k^{\text{МК}} \\ \alpha_k^{\text{СН}} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{v}_k = \begin{bmatrix} v_{\alpha,k}^{\text{МК}} \\ v_{\beta,k}^{\text{МК}} \\ v_{\gamma,k}^{\text{МК}} \\ v_{\alpha,k}^{\text{СН}} \end{bmatrix},$$

$$\mathbf{\Gamma}_y = \begin{bmatrix} \sigma_{\alpha}^{\text{МК}} & 0 & 0 & 0 \\ 0 & \sigma_{\beta}^{\text{МК}} & 0 & 0 \\ 0 & 0 & \sigma_{\gamma}^{\text{МК}} & 0 \\ 0 & 0 & 0 & \sigma_{\alpha}^{\text{СН}} \end{bmatrix},$$

где $v_{\alpha,k}^{\text{МК}}, \dots, v_{\alpha,k}^{\text{СН}}$ — стандартные независимые белые гауссовы шумы угломерных каналов ЦМК и азимутального канала МСН соответственно; $\sigma_{\alpha}^{\text{МК}}, \dots, \sigma_{\alpha}^{\text{СН}}$ — среднеквадратическая ошибка измерений в каналах, и приходим к уравнению измерения (2).

С учетом принятых обозначений соотношения (3) переписываются в виде

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + \mathbf{P}_k \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y}_k - \mathbf{H} \hat{\mathbf{x}}_{k-1}) + \mathbf{P}_k \mathbf{P}_{k/k-1}^{-1} \mathbf{D} \mathbf{w}_{k-1}, \quad \text{н. у. } \hat{\mathbf{x}}_0, \quad (9)$$

$$\mathbf{P}_{k/k-1} = \mathbf{P}_{k-1} + \mathbf{Q};$$

$$\mathbf{P}_k = (\mathbf{P}_{k/k-1}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} =$$

$$= \mathbf{P}_{k/k-1} - \mathbf{P}_{k/k-1} \mathbf{H}^T (\mathbf{H} \mathbf{P}_{k/k-1} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{P}_{k/k-1},$$

н. у. $\mathbf{P}_0$,

где

$$\mathbf{Q} = \mathbf{\Gamma} \mathbf{\Gamma}^T = \begin{bmatrix} T^2 \sigma_w^2 & 0 & 0 \\ 0 & T^2 \sigma_w^2 & 0 \\ 0 & 0 & T^2 \sigma_w^2 \end{bmatrix}$$

— ковариационная матрица (матрица дисперсий) возмущающего шума;

$$\mathbf{R} = \mathbf{\Gamma}_y \mathbf{\Gamma}_y^T = \begin{bmatrix} \sigma_{\alpha}^{2\text{МК}} & 0 & 0 & 0 \\ 0 & \sigma_{\beta}^{2\text{МК}} & 0 & 0 \\ 0 & 0 & \sigma_{\gamma}^{2\text{МК}} & 0 \\ 0 & 0 & 0 & \sigma_{\alpha}^{2\text{СН}} \end{bmatrix}$$

— ковариационная матрица измерительного шума.

После простых промежуточных преобразований

$$\mathbf{A} = \mathbf{H}^T \mathbf{R}^{-1} = \begin{bmatrix} 1/\sigma_{\alpha}^{2\text{МК}} & 0 & 0 & 1/\sigma_{\alpha}^{2\text{СН}} \\ 0 & 1/\sigma_{\beta}^{2\text{МК}} & 0 & 0 \\ 0 & 0 & 1/\sigma_{\gamma}^{2\text{МК}} & 0 \end{bmatrix};$$

$$\mathbf{B} = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} = \mathbf{A} \mathbf{H} =$$

$$= \begin{bmatrix} (1/\sigma_{\alpha}^{2\text{МК}} + 1/\sigma_{\alpha}^{2\text{СН}}) & 0 & 0 \\ 0 & 1/\sigma_{\beta}^{2\text{МК}} & 0 \\ 0 & 0 & 1/\sigma_{\gamma}^{2\text{МК}} \end{bmatrix}$$

соотношения (9) принимают окончательный вид

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + \mathbf{P}_k \mathbf{A} (\mathbf{y}_k - \mathbf{H} \hat{\mathbf{x}}_{k-1}) + \mathbf{P}_k \mathbf{P}_{k/k-1}^{-1} \mathbf{D} \mathbf{w}_{k-1}, \quad \text{н. у. } \hat{\mathbf{x}}_0; \quad (10)$$

$$\mathbf{P}_{k/k-1} = \mathbf{P}_{k-1} + \mathbf{Q}; \quad (11)$$

$$\mathbf{P}_k = (\mathbf{P}_{k/k-1}^{-1} + \mathbf{B})^{-1} =$$

$$= \mathbf{P}_{k/k-1} - \mathbf{P}_{k/k-1} \mathbf{H}^T (\mathbf{H} \mathbf{P}_{k/k-1} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{P}_{k/k-1},$$

н. у. $\mathbf{P}_0$. (12)

Выражения (10)–(12) могут служить основой моделирования и практической реализации алгоритма комплексной обработки информации угломерного канала навигационной системы подвижного объекта.

Отметим, что здесь симметрическая матрица

$$\mathbf{P}_k = \begin{bmatrix} \sigma_{\hat{\alpha},k}^2 & P_{\hat{\alpha}\hat{\beta},k} & P_{\hat{\alpha}\hat{\gamma},k} \\ P_{\hat{\beta}\hat{\alpha},k} & \sigma_{\hat{\beta},k}^2 & P_{\hat{\beta}\hat{\gamma},k} \\ P_{\hat{\gamma}\hat{\alpha},k} & P_{\hat{\gamma}\hat{\beta},k} & \sigma_{\hat{\gamma},k}^2 \end{bmatrix}$$

есть текущая ковариационная матрица ошибок оценивания, диагональные элементы которой $\sigma_{\hat{\alpha},k}^2, \sigma_{\hat{\beta},k}^2, \sigma_{\hat{\gamma},k}^2$ — дисперсии ошибок оценок $\hat{\alpha}_k, \hat{\beta}_k, \hat{\gamma}_k$, а $P_{\hat{\alpha}\hat{\beta},k} = P_{\hat{\beta}\hat{\alpha},k}$, $P_{\hat{\alpha}\hat{\gamma},k} = P_{\hat{\gamma}\hat{\alpha},k}$, $P_{\hat{\beta}\hat{\gamma},k} = P_{\hat{\gamma}\hat{\beta},k}$ — моменты взаимной корреляции (ковариации) ошибок.

Анализ компонентов алгоритма

1. Рассмотрим второе слагаемое в (10). После несложных преобразований получим

$$\mathbf{P}_k \mathbf{A}(\mathbf{y}_k - \mathbf{H} \hat{\mathbf{x}}_{k-1}) = \mathbf{P}_k \mathbf{A} \begin{bmatrix} \alpha_k^{\text{МК}} - \hat{\alpha}_{k-1} \\ \beta_k^{\text{МК}} - \hat{\beta}_{k-1} \\ \gamma_k^{\text{МК}} - \hat{\gamma}_{k-1} \\ \alpha_k^{\text{СН}} - \hat{\alpha}_{k-1} \end{bmatrix} = \mathbf{P}_k \begin{bmatrix} \frac{\alpha_k^{\text{МК}} - \hat{\alpha}_{k-1}}{\sigma_\alpha^2} + \frac{\alpha_k^{\text{СН}} - \hat{\alpha}_{k-1}}{\sigma_\alpha^2} \\ \frac{\beta_k^{\text{МК}} - \hat{\beta}_{k-1}}{\sigma_\beta^2} \\ \frac{\gamma_k^{\text{МК}} - \hat{\gamma}_{k-1}}{\sigma_\gamma^2} \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma_{\hat{\alpha},k}^2 \left(\frac{\alpha_k^{\text{МК}} - \hat{\alpha}_{k-1}}{\sigma_\alpha^2} + \frac{\alpha_k^{\text{СН}} - \hat{\alpha}_{k-1}}{\sigma_\alpha^2} \right) + P_{\hat{\alpha}\hat{\beta},k} \frac{\beta_k^{\text{МК}} - \hat{\beta}_{k-1}}{\sigma_\beta^2} + P_{\hat{\alpha}\hat{\gamma},k} \frac{\gamma_k^{\text{МК}} - \hat{\gamma}_{k-1}}{\sigma_\gamma^2} \\ P_{\hat{\beta}\hat{\alpha},k} \left(\frac{\alpha_k^{\text{МК}} - \hat{\alpha}_{k-1}}{\sigma_\alpha^2} + \frac{\alpha_k^{\text{СН}} - \hat{\alpha}_{k-1}}{\sigma_\alpha^2} \right) + \sigma_{\hat{\beta},k}^2 \frac{\beta_k^{\text{МК}} - \hat{\beta}_{k-1}}{\sigma_\beta^2} + P_{\hat{\beta}\hat{\gamma},k} \frac{\gamma_k^{\text{МК}} - \hat{\gamma}_{k-1}}{\sigma_\gamma^2} \\ P_{\hat{\gamma}\hat{\alpha},k} \left(\frac{\alpha_k^{\text{МК}} - \hat{\alpha}_{k-1}}{\sigma_\alpha^2} + \frac{\alpha_k^{\text{СН}} - \hat{\alpha}_{k-1}}{\sigma_\alpha^2} \right) + P_{\hat{\gamma}\hat{\beta},k} \frac{\beta_k^{\text{МК}} - \hat{\beta}_{k-1}}{\sigma_\beta^2} + \sigma_{\hat{\gamma},k}^2 \frac{\gamma_k^{\text{МК}} - \hat{\gamma}_{k-1}}{\sigma_\gamma^2} \end{bmatrix}. \quad (13)$$

Очевидно, что этим слагаемым обеспечивается весовая обработка невязок (разностей измеренных и оценочных значений углов), при этом веса определяются как дисперсиями ошибок измерений $\sigma_\alpha^2, \dots, \sigma_\alpha^2$, так и текущими дисперсиями $\sigma_{\hat{\alpha},k}^2, \sigma_{\hat{\beta},k}^2, \sigma_{\hat{\gamma},k}^2$ и ковариациями $P_{\hat{\alpha}\hat{\beta},k}, P_{\hat{\alpha}\hat{\gamma},k}, P_{\hat{\beta}\hat{\gamma},k}$ ошибок оценивания (последнее говорит о статической взаимосвязи оценок $\hat{\alpha}_k, \hat{\beta}_k, \hat{\gamma}_k$).

Важно подчеркнуть, что в расчете дисперсий и ковариаций ошибок оценок участвует ковариационная матрица $\mathbf{Q}$ измерительных ошибок ДУС [см. соотношение (11)]. Тем самым обеспечивается их корректный учет.

2. Анализ первых слагаемых элементов вектора (13) показывает, что при наличии двух измерений азимута $\alpha_k^{\text{МК}}, \alpha_k^{\text{СН}}$ больший вес в результирующей оценке придается измерению с меньшей погрешностью. Так, при $\sigma_\alpha^2 < \sigma_\alpha^2$ в оценке превалирует измерение $\alpha_k^{\text{СН}}$. При $\sigma_\alpha^2 \ll \sigma_\alpha^2$ вклад $\alpha_k^{\text{СН}}$ является определяющим (измерение $\alpha_k^{\text{МК}}$ в результате практически не участвует). И наоборот, при низкой точности (потере данных) азимутального канала МСН ($\sigma_\alpha^2 \gg \sigma_\alpha^2$) основной вклад в результирующую оценку дает азимутальный канал ЦМК (в пределах азимутальный канал модуля при обработке просто отсекается).

3. Особо рассмотрим третье слагаемое в (10).

Наличие члена $\mathbf{Dw}_{k-1} = [\mathbf{Tw}_{z,k-1}^{\text{ДУС}}, \mathbf{Tw}_{y,k-1}^{\text{ДУС}}, \mathbf{Tw}_{x,k-1}^{\text{ДУС}}]^T$, представляющего собой вектор парциальных приращений углов на k -м шаге, вы-

звано необходимостью восполнить в текущей оценке углов $\hat{\mathbf{x}}_{k-1}$ вычитаемый из $\mathbf{y}_k$ вектор приращений $\mathbf{Dw}_{k-2}$, содержащийся в оценке $\hat{\mathbf{x}}_{k-1}$ [см. (10) при $k-1$].

В этом состоит реализуемый в статье механизм комплексирования ДУС с другими датчиками углов. Множитель $\mathbf{P}_k \mathbf{P}_{k-1}^{-1}$ служит здесь для пересчета слагаемого $\mathbf{Dw}_{k-1}$, обеспечивающего оптимальность предсказанного значения $\hat{\mathbf{x}}_{k-1}$, в соответствующее слагаемое для оптимальной оценки $\hat{\mathbf{x}}_k$. (Заметим, что в большинстве работ по комплексной фильтрации этот момент незаслуженно опускается.)

Задание начальных условий

На практике возможны различные варианты задания начальных условий в уравнениях (10) и (12), связанные с тем или иным набором измерительных данных, привлекаемых для их формирования на первом (нескольких первых) шаге поступления и обработки измерений. Простыми и наглядными являются следующие варианты.

Вариант 1. В качестве н. у. $\hat{\mathbf{x}}_0$ выбирается первый отсчет показаний ЦМК, т. е. $\hat{\mathbf{x}}_0^{(1)} = [\alpha_1^{\text{МК}}, \beta_1^{\text{МК}}, \gamma_1^{\text{МК}}]^T$. Дальнейший счет индекса k проводится со второго измерительного шага. Очевидно, что в этом случае

$$\mathbf{P}_0^{(1)} = \begin{bmatrix} \sigma_\alpha^2 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & \sigma_\gamma^2 \end{bmatrix}.$$

Вариант 2. Этот вариант отличается от предыдущего тем, что вместо $\alpha_1^{\text{МК}}$ используется отсчет $\alpha_1^{\text{СН}}$ (объект движется), т. е.

$$\hat{\mathbf{x}}_0^{(2)} = [\alpha_1^{\text{СН}}, \beta_1^{\text{МК}}, \gamma_1^{\text{МК}}]^T;$$

$$\mathbf{P}_0^{(2)} = \begin{bmatrix} \sigma_\alpha^{2\text{СН}} & 0 & 0 \\ 0 & \sigma_\beta^{2\text{МК}} & 0 \\ 0 & 0 & \sigma_\gamma^{2\text{МК}} \end{bmatrix}.$$

Вариант 3. Для формирования н. у. используются данные ДУС. Моментом времени t_0 полагаются время начала очередного маневра поворота объекта относительно углов, соответствующих времени $t_{\text{ок}}$ окончания предыдущего маневра. Будем считать, что к моменту $t_{\text{ок}}$ обработка угловых измерений в ходе закончившегося маневра поворота обеспечивает высокую точность оценки начальных углов $\hat{\alpha}_{\text{ок}}, \hat{\beta}_{\text{ок}}, \hat{\gamma}_{\text{ок}}$. В этом случае

$$\hat{\mathbf{x}}_0^{(3)} = [\hat{\alpha}_{\text{ок}} + Tw_{z,1}^{\text{ДУС}}, \hat{\beta}_{\text{ок}} + Tw_{y,1}^{\text{ДУС}}, \hat{\gamma}_{\text{ок}} + Tw_{x,1}^{\text{ДУС}}]^T;$$

$$\mathbf{P}_0^{(3)} = \mathbf{Q} = \begin{bmatrix} T^2 \sigma_\omega^2 & 0 & 0 \\ 0 & T^2 \sigma_\omega^2 & 0 \\ 0 & 0 & T^2 \sigma_\omega^2 \end{bmatrix}.$$

Возможны другие, более сложные сочетания начальных измерений при формировании $\hat{\mathbf{x}}_0, \mathbf{P}_0$.

Заключение

Данная работа выполнена в предположении постоянства априори известных значений среднеквадратической ошибки погрешностей измерений $\sigma_w, \sigma_\alpha^{\text{МК}}, \dots, \sigma_\alpha^{\text{СН}}$. На практике требуется более детальное описание погрешностей и расчет текущих значений их среднеквадратических ошибок. Вместе с тем известные результаты теории и практики прецизионных измерений показывают, что возникающие при этом трудности в значительной мере устраняются применением к погрешностям так называемого «гарантирующего подхода» [12].

Литература

1. Ли Р. Оптимальные оценки, определение характеристик и управление. — М.: Наука, 1966. — 176 с.
2. Ярлык М. С. Статистическая теория радионавигации. — М.: Радио и связь, 1985. — 344 с.
3. Ярлык М. С., Миронов М. А. Марковская теория оценивания случайных процессов. — М.: Радио и связь, 1993. — 464 с.
4. Кузовков Н. Т., Салычев О. С. Инерциальная навигация и оптимальная фильтрация. — М.: Машиностроение, 1982. — 216 с.
5. Иванов А. В. Навигационные системы подвижных наземных объектов. Алгоритмы обработки информации в угловом канале // Радиотехника. 2013. № 4. С. 46–54.
6. Каплин А. Ю., Степанов М. Г. Использование автономной навигационной системы высокоточного позиционирования пешехода на местности // Информационно-управляющие системы. 2015. № 6. С. 86–92. doi:10.15217/issn1684-8853.2015.6.86
7. Дэвидсон П., Такала Я. Разработка алгоритмов инерциальной навигационной системы с учетом

- особенностей походки человека // Гирокоспия и навигация. 2013. № 1. С. 86–94.
8. Salychov O. S. Applied Inertial Navigation: Problems and Solutions. — М.: BMSTU Press, 2004. — 306 p.
9. Каплин А. Ю., Степанов М. Г. Модель и алгоритм комплексной обработки информации азимутального канала пешеходной навигационной системы // Изв. вузов. Приборостроение. 2016. Т. 59. № 3. С. 181–188.
10. Иванов Д. С. и др. Калибровка датчиков для определения ориентации малого космического аппарата: препринт. — ИПМ РАН, 2010. № 28. — 30 с.
11. Пат. 2572109 РФ, МПК C1 G01C 17/38. Способ калибровки электронного магнитного компаса / А. Ю. Каплин (РФ), М. Г. Степанов (РФ), А. Ю. Веревкин (РФ). — № 2014129725/28; заявл. 18.07.14; опубл. 27.12.15, Бюл. № 1. — 2 с.
12. Бахшиян Б. Ц., Назиров Р. Р., Эльясберг П. Е. Определение и коррекция движения (гарантирующий подход). — М.: Наука. Гл. ред. физ.-мат. лит., 1980. — 360 с.

UDC 621.396.96

doi:10.15217/issn1684-8853.2016.2.89

Analysis of Algorithm for Complex Processing of Goniometric Information in a Moving Object Navigation System

Kaplin A. Y.^a, PhD, Tech., a.kaplin@list.ruStepanov M. G.^a, Dr. Sc., Tech., Professor, smg099@mail.ru^aОАО «Radioavionika», 4 B, Troitskii Pr., 190005, Saint-Petersburg, Russian Federation

Introduction: For complex inertial sensors combined in mobile unit navigation systems with navigation sensors of other types (digital magnetic compass, satellite navigation module and other devices), the latter are used as an external auxiliary source of information for the selection and evaluation (using the Kalman filtering algorithm) of systematic (slowly varying) errors of inertial sensors, and their subsequent correction for an open or closed circuit. As a result, the accuracy of the navigation system of the object is determined by the accuracy of the estimates of these errors and weakly depends on the contribution of the measurements from other sensors. **Purpose:** The goal is to develop and analyze an algorithm for complex processing of goniometric information in the navigation system of a movable object. **Results:** For the azimuth channel of a movable object navigation system, an algorithm built on the basis of the Kalman filter is proposed which performs complex processing of the total volume of the measurement information. The contribution of individual components into the processing is discussed and substantiated. We have demonstrated the functioning of the weight summation mechanism implemented by the Kalman filtering algorithm for goniometric data. Options have been described for defining the filter initial conditions. **Practical relevance:** The formulated comprehensive treatment algorithm can be used as a basis for a specific software navigation system of ground-based and near-Earth mobile objects. From the methodological point of view, the results can be useful for young professionals who want to understand and master the applied aspects of Kalman filtering.

Keywords — Complex Processing, Kalman Filter, Azimuth Channel, Navigation System, Moving Object.

References

1. Li P. *Optimal'nye otsenki, opredelenie kharakteristik i upravlenie* [Optimal Assessment, Characterization and Management]. Moscow, Nauka Publ., 1966. 176 p. (In Russian).
2. Yarlykov M. S. *Statisticheskaya teoriya radionavigatsii* [Statistical Theory of Navigation]. Moscow, Radio i svyaz' Publ., 1985. 344 p. (In Russian).
3. Yarlykov M. S., Mironov M. A. *Markovskaya teoriya otsenivaniya sluchainykh protsessov* [Markov's Estimation Theory of Stochastic Processes]. Moscow, Radio i svyaz' Publ., 1993. 464 p. (In Russian).
4. Kuzovkov N. T., Salychev O. S. *Inertsial'naya navigatsiya i optimal'naya fil'tratsiya* [Inertial Navigation and Optimal Filtering]. Moscow, Mashinostroenie Publ., 1982. 216 p. (In Russian).
5. Ivanov A. V. Navigation Systems Moving Ground Targets. Algorithms of Processing of the Information in the Angular Channel. *Radiotekhnika*, 2013, no. 4, pp. 46–54 (In Russian).
6. Kaplin A. Y., Stepanov M. G. Using Autonomous Navigation System of High-precision Positioning of a Pedestrian on the Ground. *Informatsionno-upravlyaiushchie sistemy* [Information and Control Systems], 2015, no. 6, pp. 86–92 (In Russian). doi:10.15217/issn1684-8853.2015.6.86
7. Devidson P., Takala J. Development of Algorithms of Inertial Navigation System Based on Features of Human Gait. *Giroskopiya i navigatsiya*, 2013, no. 1, pp. 86–54 (In Russian).
8. Salychev O. S. *Applied Inertial Navigation: Problems and Solutions*. Moscow, BMSTU Press, 2004. 306 p.
9. Kaplin A. Y., Stepanov M. G. Model and Algorithm of Complex Information Processing Azimuth Channel Pedestrian Navigation System. *Izvestiya vuzov. Priborostroenie*, 2016, vol. 59, no. 3, pp. 181–188 (In Russian).
10. Ivanov D. S., et al. *Kalibrovka datchikov dlia opredeleniya orientatsii malogo kosmicheskogo apparata* [The Calibration of Sensors to Determine the Orientation of Small Spacecraft]. Moscow, IPM RAN Publ., 2010, no. 28. 36 p. Preprint (In Russian).
11. Kaplin A. Y., Stepanov M. G., Veryovkin A. Yu., et al. *Sposob kalibrovki elektronnoy magnitnoy kompassa* [A Method of Calibrating an Electronic Magnetic Compass]. Patent RF, no. 2572109, 2014.
12. Bakhshiyev B. Ts., Nazirov R. R., Elyasberg P. E. *Opreделение i korrektsiya dvizheniya (garantiruyushchii podkhod)* [Determination and Motion Compensation (guarantee approach)]. Moscow, Nauka Publ., 1980. 360 p. (In Russian).

СТРУКТУРИРОВАНИЕ КОНТЕНТА ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ ДАЛЬНЕЙШЕГО ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА. ПРИМЕР ФОРМИРОВАНИЯ СТРУКТУРИРОВАННОГО КОНТЕНТА УЧЕБНО-ПРОИЗВОДСТВЕННОЙ ДЕЯТЕЛЬНОСТИ

Г. Г. Куликов^а, доктор техн. наук, профессор

В. В. Антонов^а, доктор техн. наук, заведующий кафедрой

М. А. Шилина^а, канд. техн. наук, доцент

А. Р. Фахруллина^а, старший преподаватель

^аУфимский государственный авиационный технический университет, Уфа, РФ

Постановка проблемы: при разработке современных предметно-ориентированных информационно-управляющих систем возникает проблема формализации и структурирования контента для учета прослеживаемости и идентификации в течение жизненного цикла процессов, связанных между собой в рассматриваемой предметной области. **Цель:** повышение эффективности учебно-производственных процессов путем возможной интеллектуальной распределенной обработки данных в рассматриваемой предметной области за счет формализации и структурирования контента, образующегося при взаимодействии вуза и предприятия в ходе подготовки обучающихся. **Результаты:** разработана формальная структурно-параметрическая модель интеллектуальной распределенной обработки данных, которая основана на базовых положениях цикла Деминга, теории множеств и теории категорий множеств. Модель позволяет идентифицировать и проследить вертикальные и горизонтальные процессы, связанные между собой в явной форме, что дает возможность получить формализованное описание системы, отвечающее требованиям алгоритмизации. **Практическая значимость:** разработанная модель детализирована до процессных моделей в нотации BPMN, которые в дальнейшем реализуются в виде интерактивных сценариев в автоматизированных web-портальных приложениях и позволяют непрерывно обрабатывать, анализировать данные и совершенствовать учебный процесс в образовательно-производственной среде.

Ключевые слова — контент предметной области, структурирование, системная модель, программа, образовательная-производственная среда, информационная система, теория категорий, теория множеств, web-портальные приложения, PDCA, BPMN, Moodle.

Введение

В соответствии с законодательством Российской Федерации [1] образовательные организации должны сформировать информационно-образовательную среду, в том числе с использованием web-портальных и дистанционных образовательных технологий, которая позволит обеспечивать взаимодействие всех участников образовательного процесса, включая работодателей. Активное участие предприятий-работодателей предусматривает прямое воздействие на учебный процесс и позволяет будущим выпускникам еще в процессе обучения находить работу по специальности благодаря системе стажировок и грантов. Университеты в свою очередь регулярно организуют и проводят курсы повышения квалификации, переподготовку сотрудников предприятий по востребованным направлениям.

Формирование образовательно-производственной среды (ОПС) позволяет осуществлять распределенную обработку данных для непрерывного совершенствования учебного процесса на всех этапах жизненного цикла (ЖЦ) подготовки обучаемого за счет автоматизации с применением web-портальных технологий как со стороны вуза, так и со стороны работодателей. Исследования

в этой области проводятся уже довольно долгое время разными авторами [2–5 и др.]. Однако в последнее время наблюдается бурное развитие web-портальных технологий.

Существующие средства автоматизации учебно-методических процессов web-портальных технологий, включая платформы для дистанционного обучения (Blackboard, Moodle, Openmeetings и др.), представляют, прежде всего, инструментарий для создания интерактивных приложений типа Workflow. Для реализации таких приложений требуется формализовать и структурировать контент исследуемой предметной области, включая модели баз данных и баз знаний, модели бизнес-процессов и т. п.

Построение семантической модели совершенствования учебного процесса с учетом внешнего контура качества

Для управления и повышения качества бизнес-процессов стандартом ГОСТ Р ИСО 9001:2008 [6] рекомендовано применение цикла PDCA (Plan–Do–Check–Act), который иначе называют циклом Деминга (циклом Шухарта — Деминга).

В работе [7] описана семантическая модель непрерывного совершенствования учебного процесса

на уровне выпускающей кафедры, отражающая реализацию типовых учебно-методических процессов в едином информационном пространстве. Данная модель позволяет формально описать структуру этого пространства в терминах объектов и их характеристик, а также при помощи функций показать правила взаимодействия бизнес-процессов. Все функции, введенные в модель, реализуются в виде конкретных сценариев в интерактивном режиме.

Однако эта модель [7] в недостаточной степени учитывает так называемый внешний контур качества образовательного процесса в части взаимодействия вуза с предприятиями-работодателями. Исправим указанные недостатки.

Введем следующие обозначения:

$\Phi(St)$ — функция, характеризующая уровень знаний, умений и владений обучаемого St ;

R — множество целей учебного процесса, зафиксированных в соответствующих регламентирующих документах (основных профессиональных образовательных программах, учебных планах, программах практик и итоговой аттестации, рабочих программах и др.): $R = \{r_1, ..., r_d\}$;

$Z(R)$ — функция получения регламентирующей документации (учебных планов, программ практик, рабочих программ дисциплин, фондов оценочных средств и др.) на основе множества целей;

$Y(R)$ — функция стандартизации полученного опыта, на основе которого может быть сформировано множество уточненных целей;

$F = \{f_1, ..., f_i\} \in Q$ — функция, описывающая требования к уровню знаний, умений, владений (ЗУВ) у обучаемых со стороны государства;

$\Psi = \{\psi_1, ..., \psi_{i_2}\} \in M$ — функция, описывающая требования к уровню обучаемых со стороны предприятий-работодателей;

$H = \{h_1, ..., h_{i_3}\} \in M$ — функция, описывающая требования учебного управления вуза к уровню обучаемых по направлениям подготовки;

M — множество требований работодателей к специалисту;

Q — множество (совокупность) ЗУВ, удовлетворяющих требованиям государства, т. е. эталонная модель обучаемого.

Отметим, что в идеальном случае $H = F \cup \Psi$, но в реальной ситуации происходит корректировка требований вуза под изменяющиеся требования работодателей, поэтому в некоторые моменты времени это соотношение может не выполняться.

Из множества требований работодателей к специалисту можем выделить подмножество $V = \{V_1, ..., V_m\}$, которое используется в основных критериях оценок и удовлетворяет непротиворечивости вышестоящим функциям.

Аналогично из множества Q можем также выделить подмножество $X = \{X_1, ..., X_n\}$, кото-

рое используется в основных критериях оценок учебным управлением университета, и подмножество $L = \{L_1, ..., L_k\}$, которое используется в основных критериях оценок органами в сфере образования РФ. Можем ввести следующие обозначения:

$GC_u(X)$ — подмножество критериев оценки качества ЗУВ на стадиях ЖЦ подготовки обучаемого с точки зрения учебного управления университета;

$GC_r(V)$ — подмножество критериев оценки качества ЗУВ на стадиях ЖЦ подготовки обучаемого с точки зрения предприятия-работодателя;

$GC_{ra}(L)$ — подмножество критериев оценки качества ЗУВ на стадиях ЖЦ подготовки обучаемого с точки зрения органов в сфере образования.

Графическое представление модели совершенствования учебного процесса с учетом внешнего контура качества приведено на рис. 1. Следовательно, $\Phi(St) \geq \Psi$ — критерий соответствия ЗУВ обучаемого требованиям работодателей. В процессе планирования учебного процесса появляются документы, которые и определяют проведение процесса подготовки обучаемых с учетом требований предприятия-работодателя к будущему выпускнику. На данном этапе определяются порядок и график проведения учебных занятий на базе промышленного предприятия; необходимое оборудование, информационные технологии и документы представлены в виде $Z(R)$. Последующий этап — процесс обучения — регламентирован документами, сформированными на предыдущем этапе.

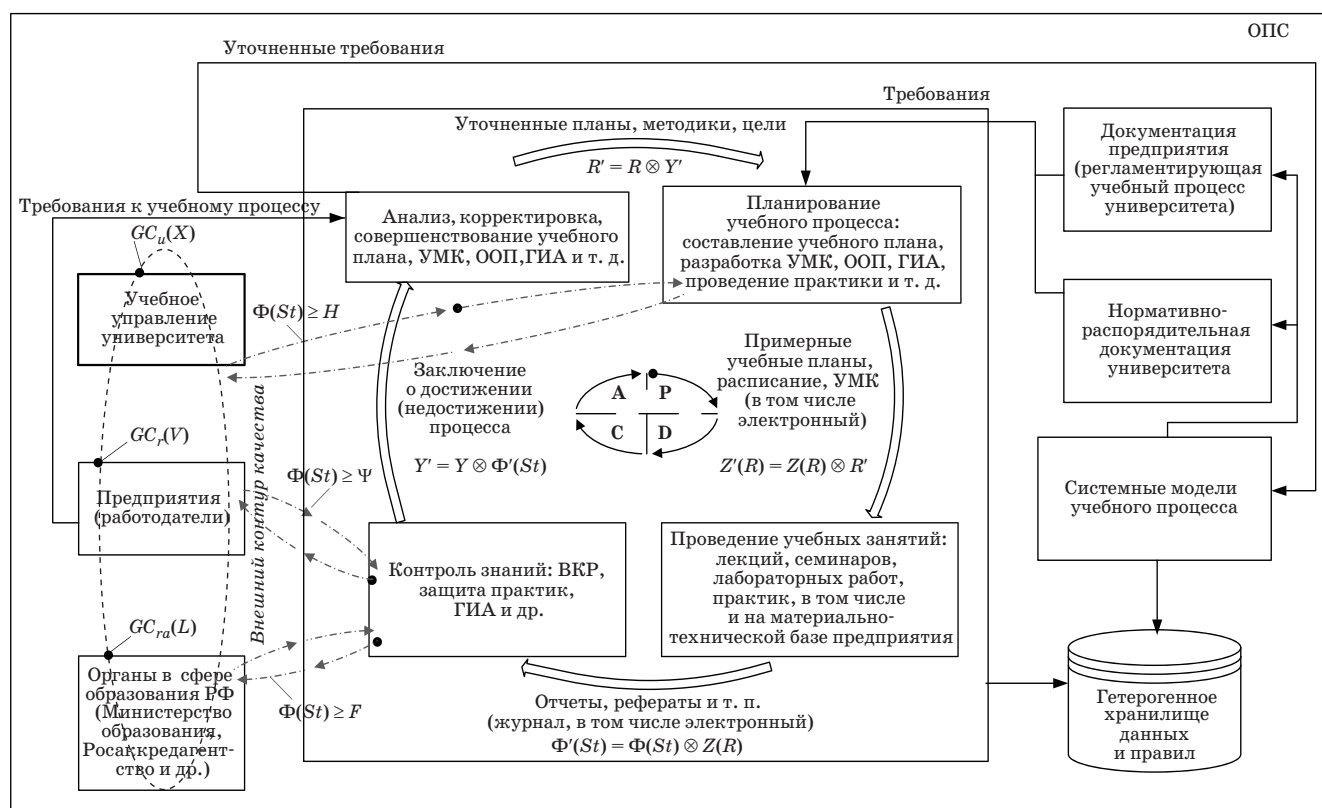
В процессе обучения имеющиеся у объекта St ЗУВ пополняются новыми значениями $\Phi'(St)$, которые определены функцией целей учебного процесса $Z(R)$. Речь идет о некотором полилинейном отображении $f: \Phi(St) \times Z(R) \rightarrow \Phi'(St)$, которое может быть рассмотрено в виде объектов некоторой категории. То есть если $f: \Phi(St) \times Z(R) \rightarrow \Phi'_1(St)$ и $g: \Phi(St) \times Z(R) \rightarrow \Phi'_2(St)$, то морфизм $f \rightarrow g$ может быть определен как гомоморфизм $h: \Phi'_1(St) \rightarrow \Phi'_2(St)$, для которого коммутативна диа-

аграмма $\Phi(St) \times Z(R) \xrightarrow{f} \Phi'_1(St) \xrightarrow{h} \Phi'_2(St) \xleftarrow{g} \Phi(St) \times Z(R)$; универсальный объект этой диаграммы есть тензорное произведение $\Phi(St)$ и $Z(R)$, т. е. имеем формулу

$$\Phi'(St) = \Phi(St) \otimes Z(R), \quad (1)$$

где знак $\otimes$ — тензорное произведение, которое отражает взаимодействие объектов по какому-либо правилу [8]. При некотором выборе базиса можем перейти к матричному исчислению.

Значение величины $\Phi'(St)$ определяется по результатам текущей, промежуточной, итоговой аттестации. Проверка ограничения $\Phi(St) \geq F$ осуществляется, например, в форме федерального



■ **Рис. 1.** Модель распределенной обработки данных совершенствования учебного процесса с учетом внешнего контура качества: УМК — учебно-методический комплекс; ООП — основная образовательная программа; ГИА — государственная итоговая аттестация; ВКР — выпускная квалификационная работа

интернет-экзамена в сфере профессионального образования, федерального интернет-экзамена для выпускников бакалавров и иных формах независимой оценки качества подготовки обучающихся на соответствие знаний требованиям федеральных образовательных стандартов высшего образования. Соответствие обучаемых в вузе требованиям предприятий-работодателей определяется в ходе прохождения ими учебных, производственных, преддипломных практик, выполнения курсовых, дипломных проектов по тематике работодателя, совместное проведение НИР, НИОКР и т. д. Учебное управление университета в соответствии с требованиями предприятий-работодателей к уровню ЗУВ обучаемых, отраженных в профессиональных стандартах и прочих регламентирующих документах, вводит изменения в ООП, учебные планы, рабочие программы дисциплин и т. п.

Вердикт о достижении или недостижении целей учебного процесса может быть сформирован на базе системы

$$\begin{cases} \Phi(St) \geq \Psi \\ \Phi(St) \geq F \end{cases} \quad (2)$$

В результате формируется множество $Y'(R) = \Phi(St)$, на основе которого может быть сформиро-

вано множество уточненных целей R' , представленных формулой

$$R' = R \otimes Y'(R). \quad (3)$$

Учитывая приведенные рассуждения, формирование новых регламентирующих документов может быть представлено формулой $Z'(R) = Z(R) \otimes R'$.

Изменения, поступающие с внешнего контура качества, например, изменение конкурентной среды предприятия-работодателя, изменение документов, регламентирующих процесс государственной аккредитации, отражаются на характеристиках учебного процесса: корректируются и совершенствуются учебные планы, рабочие программы, фонды оценочных средств, нормативно-распорядительные документы университета и т. п.

Между объектами выбранных подмножеств критериев каждого ведомства (GC_u , GC_r , GC_{ra}) могут быть установлены соответствия, реализующие сопоставление (отображение) критерия определенному критерию другого ведомства. То есть существуют X_i , V_j , L_e , при которых эти значения либо равны $GC_u(X_i) = GC_r(V_j) = GC_{ra}(L_e)$, где $i \in [1, n]$, $j \in [1, m]$, $e \in [1, k]$, либо существуют их отображения между собой. Данные подмножества образуют категории, для каждой пары объектов которых задано множество морфизмов. Связь между

объектами категорий реализуют отображения, сохраняющие структуру — функторы. Объектами в этой категории являются множества, морфизмами — отображения множеств. Можем сделать следующие выводы: приведенное подмножество критериев оценки образует класс объектов. Для любых двух объектов из данного класса (обозначим $GC_u(X_1)$ и $GC_u(X_2)$) установлено множество морфизмов $\text{Hom}(GC_u(X_1), GC_u(X_2))$, для которых определена их композиция, например:

$$\begin{aligned} g_{GC_u} &\in \text{Hom}(GC_u(X_1), GC_u(X_2)); \\ f_{GC_u} &\in \text{Hom}(GC_u(X_2), GC_u(X_3)); \\ g_{GC_u} \otimes f_{GC_u} &\in \text{Hom}(GC_u(X_1), GC_u(X_3)). \end{aligned}$$

При этом подмножество критериев оценки на стадиях ЖЦ учебного управления университета образуют категорию множеств. Проводя аналогичные рассуждения для $GC_r(V)$ и $GC_{ra}(L)$, приходим к тем же результатам. Отсюда в качестве объектов учета можем принять множество характеристик обучающихся — Stq , определив им в качестве параметров подмножества X , V и L , по значениям которых может определяться степень соответствия приведенным выше критериям, т. е.

$$Stq = \langle X, V, L \rangle. \quad (4)$$

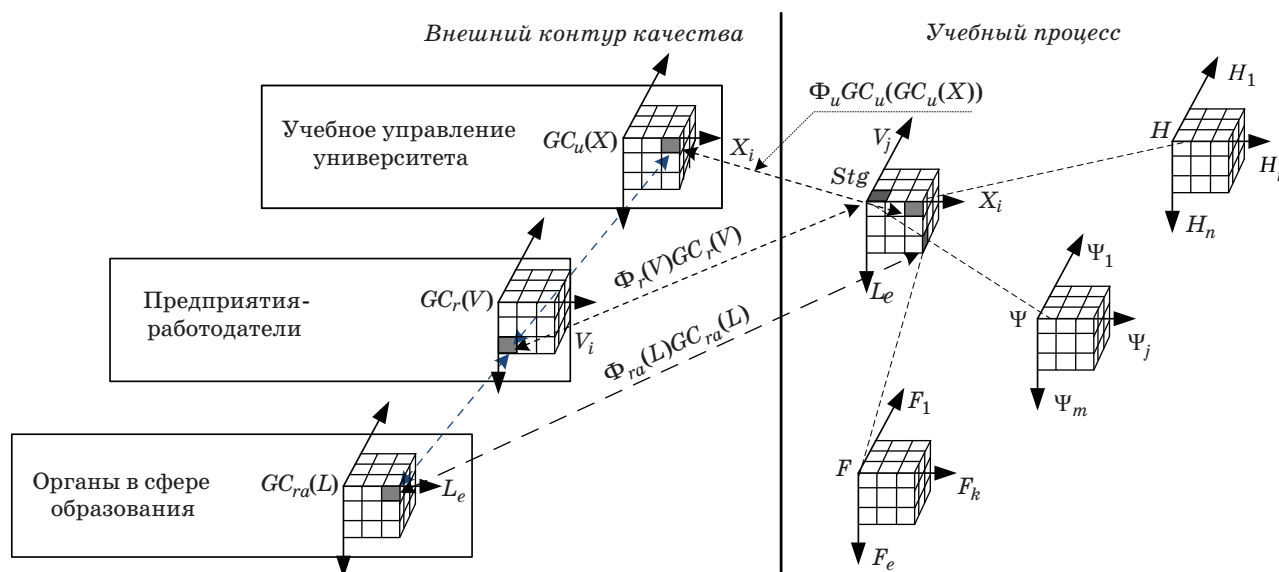
Функциональная связь с внешним контуром качества, где объекты представлены в виде многомерных матриц приведенных выше критериев: для предприятий-работодателей — $GC_r(V)$, органов в сфере образования РФ — $GC_{ra}(L)$, учебного управления университета — $GC_u(X)$ — показана на рис. 2. В результате внешний контур качества образует вертикальные связи в виде совокупно-

сти взаимосвязанных задач, направленных на достижение критериев, описанных выше, т. е. бизнес-процессов, при выполнении следующих условий (возможности совместимости ведомственных критериев):

$$\begin{aligned} GC_u(Stq)GC_{ra}(Stq)^1 &\neq \emptyset; \\ GC_u(Stq)GC_r(Stq) &\neq \emptyset. \end{aligned} \quad (5)$$

Таким образом, может быть учтена семантическая составляющая взаимодействия объектов в ОПС в течение стадий ЖЦ и применена модель единого хранилища данных для систематизации процессов сбора и обработки данных об учебном процессе.

Учитывая наличие рекуррентных соотношений внутри модели совершенствования учебного процесса и наличие описанной выше связи между объектами категорий, реализующих отображения, с сохранением структур (функторов), представляется возможным для совмещенного процесса применить принцип организации цикла Деминга, а также детализировать данные и метаданные бизнес-процессов ОПС в виде процессных моделей в нотации BPMN, которые в дальнейшем реализуются в виде интерактивных сценариев в автоматизированных web-портальных приложениях и позволяют непрерывно обрабатывать, анализировать данные и совершенствовать учебный процесс в ОПС. Наличие математического описания и формул (1)–(5) позволяет сформировать структуру web-портальных приложений рассматриваемой предметной области в виде модели математической категории множеств. Сочетание теоретико-множественного представления приведенных положений и принципа организации цикла Деминга дает возможность применить



■ Рис. 2. Графическое представление взаимодействия внешнего контура качества с учебным процессом

инкрементную модель, а также использовать для правильного отображения основных свойств данного процесса как системы математические модели реальных объектов и язык теории категорий множеств. Разработанная модель позволяет применить для анализа и прогнозирования состояний системы математический аппарат, а выделенная при этом рекурсивная зависимость (3) — организовать процесс адаптации модели на каждой стадии ЖЦ учебного процесса.

Заключение

В работе предложена модель распределенного сбора и обработки данных для интерактивного управления и совершенствования учебного про-

цесса с учетом требований внешнего контура качества. Модель построена на основе положений теории категорий множеств и в соответствии с методологией BPMN (в форме интерактивных сценариев в ОПС), а также детализирована до комплексной модели бизнес-процессов, реализуемой в web-портальных приложениях. В качестве примера web-портальных приложений используется система электронного обучения.

Работа выполнена при поддержке гранта Российского фонда фундаментальных исследований инициативных научных проектов за 2016 г., выполняемых молодыми учеными, № 16-37-00064 «Программное обеспечение для многоуровневого структурирования контента информационного пространства по системной модели».

Литература

1. Федеральный закон «Об образовании в Российской Федерации» от 29 дек. 2012 г. № 273 (ред. от 31 декабря 2014 г.).
2. Мартынов В. В., Рыков В. И., Закиева Е. Ш. Управление функциональной структурой и контентом в обучающих системах // Вестник Уфимского государственного авиационного технического университета. 2010. Т. 14. № 5(40). С. 157–162.
3. Федоров И. Г. и др. Особенности проектирования системы управления бизнес-процессами на примере электронного вуза/ И. Г. Федоров, К. С. Курышев, А. В. Данилов, Е. А. Завражная, В. И. Швей // Открытое образование. МЭСИ. 2012. № 4. С. 75–81.
4. Васильев В. И. Интеллектуальные системы управления. Теория и практика. — М.: Радиотехника, 2009. — 392 с.
5. Воробьева М. С. Построение модели интеграции данных в информационно-управляющих системах // Модернизация образования в условиях глобализации: тр. Междунар. конф., Тюмень, 14–15 сентября 2005 г. Тюмень, 2005. С. 26–28.
6. ГОСТ Р ИСО–9000:2008. Системы менеджмента качества. Основные положения и словарь. — М.: Изд-во стандартов, 2009. — 61 с.
7. Куликов Г. Г., Антонов В. В., Старцев Г. В., Шилина М. А. Поддержка управления учебным процессом на основе информационных технологий многомерного анализа данных (на примере высшего учебного заведения). — Уфа: УГАТУ, 2013. — 152 с.
8. Куликов Г. Г., Михеев А. Г. Особенности реализации процессного подхода и обучения управлению бизнес-процессами при помощи свободного программного обеспечения с открытым кодом // Открытое образование. МЭСИ. 2011. № 4. С. 47–57.

UDC 004:04

doi:10.15217/issn1684-8853.2016.2.95

Structuring Domain Content for Further Data Mining: an example of Forming Structured Content for Education-and-Production Activity

Kulikov G. G.^a, Dr. Sc., Tech., Professor, gennadyg_98@Yahoo.com

Antonov V. V.^a, Dr. Sc., Tech., Head of Chair, boss@bashkortostan.ru

Shilina M. A.^a, PhD, Tech., Associate Professor, maria.shilina@gmail.com

Fakhrullina A. R.^a, Senior Lecturer, almirafax@mail.ru

^aUfa State Aviation Technical University, 12, Karla Marksa St., 450000, Ufa, Russian Federation

Intriduction: Developing modern object-oriented information management systems often has the problem of formalizing and structuring the content in order to take into account the lifecycle traceability and identification of interlinked processes from the subject area. **Purpose:** The effectiveness of training and production processes should be improved by the intellectual potential of distributed data processing in the subject area through the formalization and structuring of the content generated during the interaction between the University and enterprises in the process of education. **Results:** A formal structural-parametric model of intelligent distributed data processing has been developed which allows you to identify and trace the vertical and horizontal processes associated with each other in an explicit form. The model is based on the basic provisions of the set theory, the category theory and the Deming cycle. It provides a formalized description of the system that meets the algorithmization requirements. **Practical relevance:** The developed

model is detailed down to process models in BPMN notation which are realized further in the form of interactive scenarios in automated web-portal applications, allowing you to continuously process and analyze the data, improving the educational process in the education-and-production environment.

Keywords — Domain Content, Structuring, System Model, Program, Education-and-Production Environment, Information System, Category Theory, Set Theory, Web-Portal Applications, PDCA, BPMN, Moodle.

References

1. Federal Law "On Education in Russian Federation" 29.12.2012 no. 273 (red. 31/12/2014 g.). (In Russian).
2. Martynov V. V., Rykov V. I., Zakieva E. Sh. Management of Functional Structure and Content in Learning Systems. *Vestnik Ufimskogo gosudarstvennogo aviatsionnogo tekhnicheskogo universiteta*, 2010, vol. 14, no. 5(40), pp. 157–162 (In Russian).
3. Fedorov I. G., Kuryshv K. S., Danilov A. V., Zavrzhnaya E. A., Shvej V. I. Features Design Business Process Management System on the Example of E-university. *Otkrytoe obrazovanie*, 2012, no. 4, pp. 75–81 (In Russian).
4. Vasilev V. I. *Intellektual'nye sistemy upravleniya. Teoriya i praktika* [Intelligent Control Systems. Theory and Practice]. Moscow, Radiotekhnika Publ., 2009. 392 p. (In Russian).
5. Vorobeva M. S. Construction of the Model of Data Integration in Information and Control Systems. *Trudy Mezhduarodnoi konferentsii "Modernizatsiya obrazovaniya v usloviakh globalizatsii"* [Proc. Int. Conf. "Modernization of Education in the Context of Globalization"]. Tyumen, 2005, pp. 26–28 (In Russian).
6. State Standard R ISO 9000: 2008. Quality Management Systems. Fundamentals and Vocabulary. Moscow, Standartov Publ., 2009. 61 p. (In Russian).
7. Kulikov G. G., Antonov V. V., Startsev G. V., Shilina M. A. *Podderzhka upravleniya uchebnym protsessom na osnove informatsionnykh tekhnologii mnogomernogo analiza dannykh (na primere vysshego uchebnogo zavedeniya)* [Educational Process Support Management based on Information Technology of Multidimensional Data Analysis (on the Example of the University)]. Ufa, UGATU Publ., 2013. 152 p. (In Russian).
8. Kulikov G. G., Mikheev A. G. Features of the Learning Process Approach and Business Process Management using Open Source Software. *Otkrytoe obrazovanie*, 2011, no. 4, pp. 47–57 (In Russian).

Научный журнал «ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ» выходит каждые два месяца.

Стоимость годовой подписки (6 номеров) для подписчиков России — 4800 рублей, для подписчиков стран СНГ — 5400 рублей, включая НДС 18%, таможенные и почтовые расходы.

Подписку на печатную версию журнала можно оформить в любом отделении связи по каталогу:

«Роспечать»: № 48060 — годовой индекс, № 15385 — полугодовой индекс,

а также через посредство подписных агентств:

«Северо-Западное агентство „Прессинформ“»

Санкт-Петербург, тел.: (812) 335-97-51, 337-23-05,

эл. почта: press@crp.spb.ru, zajavka@crp.spb.ru,

сайт: <http://www.pinform.spb.ru>

«МК-Периодика» (РФ + 90 стран)

Москва, тел.: (495) 681-91-37, 681-87-47,

эл. почта: export@periodicals.ru, сайт: <http://www.periodicals.ru>

«Информнаука» (РФ + ближнее и дальнее зарубежье)

Москва, тел.: (495) 787-38-73, эл. почта: informnauka3@yandex.ru,

сайт: <http://www.informnauka.com>

«Деловая пресса»

Москва, тел.: (495) 962-11-11, эл. почта: podpiska@delpress.ru,

сайт: <http://delpress.ru/contacts.html>

«Коммерсант-Курьер»

Казань, тел.: (843) 291-09-99, 291-09-47, эл. почта: kazan@komcur.ru,

сайт: <http://www.komcur.ru/contacts/kazan/>

«Урал-Пресс» (филиалы в 40 городах РФ)

Сайт: <http://www.ural-press.ru>

«Идея» (Украина)

Сайт: <http://idea.com.ua>

«BTL» (Узбекистан)

Сайт: <http://btl.sk.uz/ru/cat17.html> и др.

На электронную версию нашего журнала (все выпуски, годовая подписка, один выпуск, одна статья)

вы можете подписаться на сайтах НЭБ: <http://elibrary.ru>;

РУКОНТ: <http://www.rucont.ru>; ИВИС: <http://www.ivis.ru/>

Полнотекстовые версии журнала за 2002–2014 гг.

в свободном доступе на сайте журнала (<http://www.i-us.ru>),

НЭБ (<http://www.elibrary.ru>)

и Киберленинки (<http://cyberleninka.ru/>

journal/n/informatsionno-upravlyayuschiesistemy).

МАТРИЦЫ ПРОПУС 92 И 116

Н. А. Балонин^а, доктор техн. наук, профессор

М. Б. Сергеев^а, доктор техн. наук, профессор

^аСанкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ

Цель: привести результаты численного поиска симметричных матриц семейства Адамара высокой размерности в форме Пропус — специализированных массивов Вильямсона или Гетхальса — Зейделя с двумя равными друг другу блоками. **Методы:** поиск матриц глобального и локального максимумов детерминанта итерационной вычислительной процедурой, ориентированной на минимизацию максимального абсолютного значения элементов ортогональной матрицы. **Результаты:** найдены новые симметричные матрица Адамара порядка 92 и взвешенная матрица $W(116, 114)$ в форме Пропус. Приведены подтверждения существования симметричных конструкций на указанных порядках с приведением конкретных реализаций. Теория матриц Адамара дополнена конструкцией Пропус. **Практическая значимость:** матрицы Адамара ортогональны по строкам и столбцам и имеют непосредственное практическое значение для задач помехоустойчивого кодирования, сжатия и маскирования видеoinформации.

Ключевые слова — симметричные матрицы, матрицы Адамара, матрицы Вильямсона, массив Гетхальса — Зейделя, массив Вайтмана — Себерри, матрицы Пропус, критские матрицы.

В задачах обработки информации матричные конструкции находят широкое применение. Наиболее часто используются матрицы Адамара (**H**) порядка n с элементами $\{1, -1\}$, для которых характерно [1] $H^T H = nI$, где I — единичные матрицы, получаемые разными способами и имеющие различные конструктивные особенности. Наибольший интерес представляют симметричные матрицы Адамара [2], методы их получения [3, 4] и обобщения на нечетные порядки [5, 6] (критские матрицы).

В настоящей работе в рассмотрение вводится матрица Пропус (**P**), являющаяся разновидностью массива Вильямсона (**H**) и получаемая из него при равенстве пары блоков $B=C$ перестановками таким образом:

$$H = \begin{pmatrix} A & B & C & D \\ -B & A & -D & C \\ -C & D & A & -B \\ -D & -C & B & A \end{pmatrix} \rightarrow$$

$$\rightarrow P = \begin{pmatrix} A & B=C & C=B & D \\ C & D & -A & -B \\ B & -A & -D & C \\ D & -C & B & -A \end{pmatrix}.$$

Здесь **A**, **B**, **C**, **D** — матрицы Вильямсона [7]. Содержательная сторона предложения выделять матрицы Пропус состоит в том, что это — **симметричные** матрицы, стоящие между бициклическими формами (матрицами из двух блоков **A**, **B**) [2, 4] и более крупными четырехблочными массивами [1, 7], которые не всегда существуют [8].

Симметрируя три последние циклические клетки с помощью реверсивной единичной матрицы **R**, можно найти больше симметричных решений с единственным симметричным блоком **A**. Обобщенная матрица Пропус (**P***) наследует черты массива Гетхальса — Зейделя [1] и имеет вид

$$P^* = \begin{pmatrix} A & BR & CR & DR \\ CR & D^T R & -A & -B^T R \\ BR & -A & -D^T R & C^T R \\ DR & -C^T R & B^T R & -A \end{pmatrix}.$$

Конструкции, подчеркивающие кососимметрию, называются массивами Вильямсона — Себерри [9, 10], симметричная матрица Пропус — симметричным массивом Балонина — Себерри [11, 12]. Название *Пропус* впервые возникло на страницах каталога [3] при изучении симметрии квазиортогональных матриц [5, 6].

Семейство симметричных матриц Пропус распадается, в свою очередь, на парные матрицы Пропус с $A=D$ порядков Сильвестра 2^k (тип GOOD), где k — натуральное число, и почти парные матрицы Пропус (тип GOODD), когда равенство $A=D$ нарушается только противоположными знаками элементов диагонали. Таковы почти все матрицы Пропус порядков не выше первой сотни, но есть и исключения (табл. 1).

Приведенные параметры показывают, что матрица Пропус — это почти бициклическая матрица, так как матрицы **A** и **D** мало отличаются друг от друга, что существенно упрощает их поиск. Взвешенные матрицы $W(n, n-2)$ [1] сосуществуют с симметричными матрицами Адамара

■ **Таблица 1.** Параметры циклических матриц A , $B=C, D$

Порядок n	Тип	Последовательности
4	GOOD	$a = [-1]; b = [1]; d = [-1];$
8	GOOD	$a = [1, -1]; b = [1, 1]; d = [1, -1];$
12	GOODD	$a = [1, 1, 1]; b = [-1, 1, 1];$ $d = [-1, 1, 1];$
16	GOOD	$a = [-1, 1, 1, 1]; b = [-1, 1, 1, 1];$ $d = [-1, 1, 1, 1];$
20	GOODD	$a = [1, 1, -1, -1, 1]; b = [-1, 1, 1, 1, 1];$ $d = [-1, 1, -1, -1, 1];$
24	GOODD $W(24, 22)$	$a = [1, 1, 1, -1, 1, 1];$ $b = [0, -1, 1, 1, -1];$ $d = [-1, 1, 1, -1, 1];$
28	GOODD	$a = [1, -1, 1, -1, -1, 1, -1];$ $b = [1, 1, 1, -1, -1, 1, 1];$ $d = [-1, -1, 1, -1, -1, 1, -1];$
32	GOOD	$a = [1, -1, 1, 1, 1, 1, 1, -1];$ $b = [-1, -1, 1, 1, -1, 1, 1, -1];$ $d = [1, -1, 1, 1, 1, 1, 1, -1];$

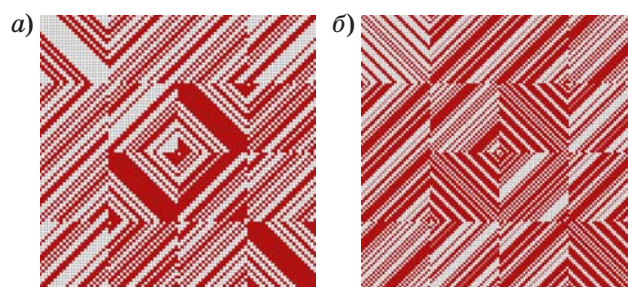
и заменяют их на порядках, для которых симметричных форм с $\mathbf{V}=\mathbf{C}$ и элементами 1, -1 нет. Цель короткого сообщения — представить две найденные редкие матрицы Пропуск порядков 92, 116.

Для нахождения Пропус необходимо вычислить три циклические матрицы, а не четыре. Этот выигрыш позволил с помощью высокопроизводительных компьютеров в 2015 г. доказать существование симметричных матриц Адамара порядков 116 и 172 [12], в справочнике [1] 2007 г. они отмечены как неизученные на предмет симметрии.

Опубликованный нами ранее оригинальный алгоритм [6, 13] позволил найти независимую версию симметричной матрицы Адамара — матрицу Пропус порядка 92, а на порядке 116 дополнил результат работы [12] взвешенной матрицей Пропус $W(116, 114)$. Алгоритм реализован на обычном компьютере, что характеризует его по-

■ **Таблица 2.** Параметры новых циклических матриц $A, B=C, D$

Порядок n	Последовательности
92	$a = [1, 1, 1, 1, 1, 1, -1, -1, 1, -1, -1, 1, 1, -1, -1, 1, -1, -1, 1, 1, 1, 1];$ $b = [1, -1, 1, 1, -1, 1, -1, -1, 1, 1, -1, 1, 1, -1, 1, 1, -1, -1, 1, 1, 1, -1];$ $d = [-1, -1, 1, 1, 1, 1, 1, 1, -1, 1, 1, 1, -1, -1, -1, 1, -1, 1, -1, 1, 1, -1];$
116	$a = [1, -1, 1, 1, -1, 1, 1, 1, -1, -1, 1, 1, 1, 1, -1, -1, 1, 1, 1, -1, -1, 1, 1, -1];$ $b = [0, 1, 1, -1, -1, 1, 1, -1, 1, -1, -1, -1, -1, -1, -1, -1, 1, 1, 1, -1, -1, 1, 1, -1];$ $d = [-1, 1, -1, 1, 1, 1, 1, -1, 1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 1, -1, 1, 1, -1];$



■ Матрицы Пропуск: матрица $\mathbf{H}_{92}$ (а) и взвешенная матрица $\mathbf{W}(116, 114)$ (б)

исковые качества. Параметры вычисленных матриц приведены в табл. 2.

Для сравнения приведем найденные матрицы Пропус — Адамара 92 (рисунок, *а*) и взвешенную $W(116, 114)$ (рисунок, *б*) симметричные матрицы.

В заключение отметим, что конструкция матрицы Пропус позволяет найти близкие аппроксимации неизвестных пока матриц Адамара порядков 668, 716, 892 и т. п. [5, 6]. Кроме ценности самого по себе данного результата, очевидно, что матрицы Пропус с их тремя клетками представляют собой необходимый элемент теории бициклических и тетрациклических форм, являясь переходным и удобным для вычислений звеном между ними, — трициклом.

Литература

1. **Craigen R., Kharaghani H.** Hadamard Matrices and Hadamard Designs // Handbook of Combinatorial Designs / Ed. by Charles J. Colbourn and Jeffrey H. Dinitz. Second ed. — Boca Raton, FL: Chapman & Hall / CRC, 2006. Pp. 273–280.
2. **Balonin N. A., Djokovic D. Z.** Symmetry of Two Circulant Hadamard Matrices and Periodic Golay Pairs // Информационно-управляющие системы. 2015. № 3(76). С. 2–16. doi:10.15217/issn1684-8853.2015.3.2

3. Balonin N. A., Djokovic D. Z., Seberry J., Sergeev M. B. Hadamard-Type Matrices. <http://math-scinet.ru/catalogue/index.php> (дата обращения: 15.11.2015).
4. Balonin N. A., Djokovic D. Z. Negaperiodic Golay Pairs and Hadamard Matrices // Информационно-управляющие системы. 2015. № 5(78). С. 2–17. doi:10.15217/issn1684-8853.2015.5.2
5. Балонин Н. А., Сергеев М. Б. Матрицы Мерсенна и Адамара // Информационно-управляющие системы. 2016. № 1(80). С. 2–15. doi:10.15217/issn1684-8853.2016.1.2

6. Балонин Н. А., Сергеев М. Б. Матрицы локального максимума детерминанта // Информационно-управляющие системы. 2014. № 1(68). С. 2–15.
7. Holzmamm W. H., Kharaghani H., Tayfeh-Rezaie B. Williamson Matrices up to Order 59// Designs, Codes and Cryptography. 2008. Vol. 46. Iss. 3. P. 343–352.
8. Djokovic D. Z. Williamson Matrices of Order $4n$ for $n=33;35;39$ // Discrete Math. 1993. Vol. 115. P. 267–271.
9. Awyzio G., and Seberry J. On Good Matrices and Skew Hadamard Matrices. http://www.uow.edu.au/~jennie/WEB/WEB15/2015_11_Good_matrices.pdf (дата обращения: 25.09.2014).
10. Seberry W. J. A Skew-Hadamard Matrix of Order 92 // Bulletin of the Australian Mathematical Society. 1971. Vol. 5. P. 203–204.
11. Balonin N. A., Seberry J. The Propus Construction for Symmetric Hadamard Matrices. 2015. <http://arxiv.org/abs/1512.01732> (дата обращения: 22.10.2015).
12. Di Matteo O., Djokovic D. Z., Kotsireas I. S. Symmetric Hadamard Matrices of Order 116 and 172 Exist // Special Matrices. 2015. Vol. 3.1. P. 227–234.
13. Балонин Н. А., Сергеев М. Б. О значении матриц начального приближения в алгоритме поиска обобщенных взвешенных матриц глобального и локального максимума детерминанта // Информационно-управляющие системы. 2015. № 6(79). С. 2–9. doi:10.15217/issn1684-8853.2015.6.2

UDC 519.614

doi:10.15217/issn1684-8853.2016.2.101

Propus Matrices 92 and 116

Balonin N. A.^a, Dr. Sc., Tech., Professor, korbendfs@mail.ru

Sergeev M. B.^a, Dr. Sc., Tech., Professor, mbse@mail.ru

^aSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

Purpose: The paper presents our results in numerical search for symmetric high-order propus-Hadamard matrices: special Williamson or Goethal–Seidel arrays with two equal blocks. **Methods:** Search for global or local maximum determinant matrices is performed via an iterative computational procedure focused on the minimization of the maximum absolute values of the elements of an orthogonal matrix. **Results:** A new symmetric Hadamard matrix of order 92 has been found, along with a symmetric weighing matrix $W(116,114)$ in Propus construction. A confirmation is given for the existence of symmetric constructions on these orders with specific implementations. The theory of Hadamard matrices is supplemented by the Propus construction. **Practical relevance:** Hadamard matrices are orthogonal by rows and columns. They have a direct practical value for the problems of error-correcting coding, video compression and masking.

Keywords — Symmetric Matrices, Hadamard Matrices, Williamson Matrices, Goethal–Seidel Array, Whiteman–Seberry Array, Propus Matrices, Cretan Matrices.

References

1. Craigen R. and Kharaghani H. Hadamard Matrices and Hadamard Designs. In: *Handbook of Combinatorial Designs*. Ed. by Charles J. Colbourn and Jeffrey H. Dinitz. Second ed. Boca Raton, FL, Chapman & Hall/CRC, 2006, pp. 273–280.
2. Balonin N. A., Djokovic D. Z. Symmetry of Two Circulant Hadamard Matrices and Periodic Golay Pairs. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 3(76), pp. 2–16 (In Russian). doi:10.15217/issn1684-8853.2015.3.2
3. Balonin N. A., Djokovic D. Z., Seberry J., Sergeev M. B. *Hadamard-Type Matrices*. Available at: <http://mathscinet.ru/catalogue/index.php> (accessed 15 November 2015).
4. Balonin N. A., Djokovic D. Z. Negaperiodic Golay Pairs and Hadamard Matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 5(78), pp. 2–17. doi:10.15217/issn1684-8853.2015.5.2
5. Balonin N. A., Sergeev M. B. Mersenne and Hadamard Matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2016, no. 1(80), pp. 2–15 (In Russian). doi:10.15217/issn1684-8853.2016.1.2
6. Balonin N. A., Sergeev M. B. Local Maximum Determinant Matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2014, no. 1(68), pp. 2–15 (In Russian).
7. Holzmamm W. H., Kharaghani H., Tayfeh-Rezaie B. Williamson Matrices up to Order 59. *Designs, Codes and Cryptography*, 2008, vol. 46, iss. 3, pp. 343–352.
8. Djokovic D. Z. Williamson Matrices of Order $4n$ for $n=33;35;39$. *Discrete Math.*, 1993, no. 115, pp. 267–271.
9. Awyzio Gene, and Seberry Jennifer. *On Good Matrices and Skew Hadamard Matrices*. Available at: http://www.uow.edu.au/~jennie/WEB/WEB15/2015_11_Good_matrices.pdf (accessed 25 September 2014).
10. Seberry Wallis Jennifer. A Skew-Hadamard Matrix of Order 92. *Bulletin of the Australian Mathematical Society*, 1971, vol. 5, pp. 203–204.
11. Balonin N. A., Seberry J. *The Propus Construction for Symmetric Hadamard Matrices*, 2015. Available at: <http://arxiv.org/abs/1512.01732> (accessed 22 October 2015).
12. Di Matteo O., Djokovic D. Z., Kotsireas I. S. Symmetric Hadamard Matrices of Order 116 and 172 Exist. *Special Matrices*, 2015, vol. 3.1, pp. 227–234.
13. Balonin N. A., Sergeev M. B. Initial Approximation Matrices in Search for Generalized Weighted Matrices of Global or Local Maximum Determinant. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 6(79), pp. 2–9 (In Russian). doi:10.15217/issn1684-8853.2015.6.2

АНАЛИЗ ЛИНЕЙНЫХ СИСТЕМ НА ОСНОВЕ ПЕРЕХОДНЫХ ХАРАКТЕРИСТИК

С. И. Зиятдинов^а, доктор техн. наук, профессор

^аСанкт-Петербургский государственный университет аэрокосмического приборостроения,
Санкт-Петербург, РФ

Введение: переход от непрерывных линейных систем к дискретным с использованием отсчетов импульсных характеристик непрерывных линейных систем-аналогов приводит к ошибкам в формировании переходных процессов. Вместе с тем динамические свойства линейных систем в полном объеме определяются их переходными характеристиками. Синтез дискретных линейных систем на базе переходных характеристик позволяет обеспечить динамические свойства дискретных линейных систем, совпадающие с динамическими свойствами непрерывных линейных систем.

Цель: разработка методики исследования линейных систем с использованием переходных характеристик. **Методы:** представление входного сигнала в виде суммы сдвинутых относительно друг друга ступенчатых воздействий и нахождение с использованием переходной характеристики выходящего сигнала линейной системы в виде суммы реакций системы на каждое ступенчатое воздействие. **Результаты:** получена новая форма интеграла наложения, позволяющая при заданном входном сигнале и известной переходной характеристике линейной системы найти ее выходной сигнал. Выдвинутые теоретические положения подтверждены конкретными примерами. **Практическая значимость:** полученная методика позволяет синтезировать дискретные линейные системы, динамические свойства которых совпадают с динамическими свойствами непрерывных линейных систем-аналогов.

Ключевые слова — линейная система, переходная характеристика, интеграл наложения.

Введение

В настоящее время теория исследования прохождения сигнала через линейную систему достаточно хорошо разработана [1–3].

Исчерпывающей характеристикой линейной системы является ее частотно-передаточная функция. Для анализа линейных систем используются два основных метода — частотный и временной. В случае частотного метода исследования линейной системы с частотной передаточной функцией $W(j\omega)$ выходной сигнал $y(t)$ при известной спектральной плотности $x(j\omega)$ входного сигнала $x(t)$ определяется на основании обратного преобразования Фурье [2]:

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(j\omega) x(j\omega) e^{j\omega t} d\omega.$$

При использовании временного метода анализа линейной системы с импульсной характеристикой $h(t)$ используется интеграл наложения (интеграл Дюамеля) [1]

$$y(t) = \int_0^{\infty} x(t - \tau) h(\tau) d\tau. \quad (1)$$

Некоторые линейные системы могут описываться дифференциальными уравнениями вида [2]

$$\begin{aligned} y^{(n)} + b_1 y^{(n-1)} + \dots + b_n y = \\ = a_0 x^{(m)} + a_1 x^{(m-1)} + \dots + a_m x, \quad m \leq n. \end{aligned}$$

Дифференциальные модели не выше второго порядка соответствуют многим реальным динамическим звеньям. Достоинство дифференциальных моделей заключается в простоте перехода описания систем из пространства времени к описанию в пространстве частоты.

Однако не всякая линейная система описывается дифференциальным уравнением. Вместе с тем в работе [1] показано, что при переходе от непрерывных линейных систем к дискретным системам существующая методика расчетов коэффициентов разностных уравнений на основе отсчетов импульсных характеристик не обеспечивает правильной реализации переходных процессов. При этом методика расчетов коэффициентов с использованием отсчетов переходных характеристик позволяет получить динамические свойства дискретных линейных систем, совпадающие с динамическими свойствами непрерывных линейных систем как в переходном, так и в установившемся режимах.

В этих условиях встает актуальная задача разработки методики анализа линейных систем на основе их переходных характеристик. Решение данной задачи составляет основное содержание настоящей работы.

Интеграл наложения с использованием переходных характеристик

Для решения поставленной задачи входной сигнал $x(t)$ линейной системы, показанный на рисунке, представим в виде суммы последовательно-

сти ступенчатых функций, начинающихся в моменты времени $t_1, t_2, \dots, t_n$:

$$\begin{aligned} x(t) &= [x(t-t_1) - x(t-t_0)] + [x(t-t_2) - x(t-t_1)] + \dots \\ &\quad \dots + [x(t-t_i) - x(t-t_{i-1})] + \dots = \\ &= \Delta x(t_1) + \Delta x(t_2) + \dots + \Delta x(t_i) + \dots, \end{aligned}$$

где $\Delta x(t_i) = x(t-t_i) - x(t-t_{i-1})$ — элементарная ступенчатая функция, возникающая в момент времени $t = t_i$.

При этом выходной сигнал линейной системы будет определяться соотношением

$$\begin{aligned} y(t) &= \Delta x(t_1)g(t-t_1) + \Delta x(t_2)g(t-t_2) + \dots \\ &\quad \dots + \Delta x(t_i)g(t-t_i) + \dots = \sum_{i=1}^{\infty} \Delta x(t_i)g(t-t_i), \end{aligned} \quad (2)$$

где $g(t)$ — переходная характеристика, являющаяся реакцией линейной системы на ступенчатое входное воздействие $\Delta x(t)$.

Пусть отсчеты входного сигнала при формировании ступенчатых функций берутся через одинаковые отрезки времени, т. е. $t_i = i\Delta T$, где ΔT — период следования отсчетов $x(t_i)$.

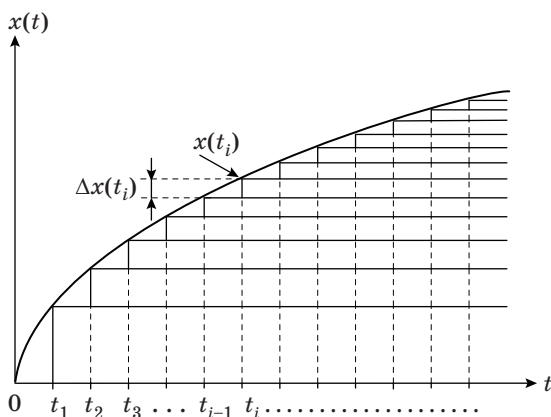
Помножив и поделив (2) на ΔT , получим

$$y(t) = \Delta T \sum_{i=1}^{\infty} \frac{\Delta x(t_i)}{\Delta T} g(t-t_i).$$

В предельном переходе при $\Delta T \rightarrow 0$ можно записать, что

$$y(t) = \int_0^t \frac{dx(\tau)}{d\tau} g(t-\tau) d\tau. \quad (3)$$

Полученное выражение (3) для выходного сигнала линейной системы является интегралом наложения, связывающим производную входного сигнала $\frac{dx(t)}{dt}$ и переходную характеристику линейной системы $g(t)$.



■ Представление сигнала в виде суммы сдвинутых ступенчатых функций

Известно [2], что импульсная характеристика линейной системы является производной от переходной характеристики $h(t) = \frac{dg(t)}{dt}$. Тогда выражение для выходного сигнала линейной системы (1) принимает вид

$$y(t) = \int_0^t x(t-\tau) \frac{dg(\tau)}{d\tau} d\tau. \quad (4)$$

Таким образом, пара равноценных преобразований (3) и (4) позволяет найти выходной сигнал линейной системы при известных входном сигнале $x(t)$ и переходной характеристике $g(t)$.

Для проверки правильности полученных результатов рассмотрим конкретные примеры.

Пример 1. Пусть входной сигнал линейной системы является единичным ступенчатым воздействием $x(t) = 1(t)$.

Для данного сигнала $\frac{dx(t)}{dt} = \delta[.]$ — дельта-функция, после подстановки которой в (3) получим

$$y(t) = \int_0^t \delta[0]g(t-\tau) d\tau = g(t).$$

В результате при единичном ступенчатом входном воздействии выходной сигнал линейной системы представляет ее переходную характеристику.

Пример 2. Рассмотрим линейную систему в виде фильтра верхних частот (ФВЧ) первого порядка, на входе которого действует гармоническое колебание вида $x(t) = \sin(\omega t)$, где ω — частота входного сигнала.

Для рассматриваемого ФВЧ переходная характеристика записывается следующим образом [2]:

$$g(t) = e^{-\omega_{cp} t},$$

где ω_{cp} — частота среза фильтра.

При этом производная входного сигнала определяется выражением

$$\frac{dx(t)}{dt} = \omega \cos(\omega t).$$

После подстановки производной входного сигнала в (3) получим

$$y(t) = \int_0^{\infty} \omega \cos(\omega \tau) e^{-\omega_{cp}(t-\tau)} d\tau.$$

Интеграл в данной формуле является табличным. В окончательном виде для установившегося режима можно записать следующее выражение для выходного сигнала ФВЧ:

$$y(t) = \frac{\omega}{\sqrt{1 + \left(\frac{\omega}{\omega_{cp}}\right)^2}} \sin\left(\omega t + \arctg \frac{\omega_{cp}}{\omega}\right).$$

Данное соотношение строго определяет выходной сигнал ФВЧ первого порядка при гармоническом входном сигнале.

Полученные в работе соотношения (3) и (4) для интеграла наложения дополняют известную методику анализа линейных систем.

Заключение

В результате проведенных исследований получила дополнительное развитие теория линейных систем. Представлена новая форма интеграла наложения, позволяющая либо при заданной производной входного сигнала и известной переходной характеристике линейной системы, либо при заданном входном сигнале и известной производной переходной характеристике найти выходной сигнал линейной системы. Разрабо-

танная методика может быть использована при синтезе по заданной переходной характеристике дискретных линейных систем, динамические свойства которых совпадают с динамическими свойствами непрерывных линейных систем-аналогов.

Литература

1. Воробьев С. Н. Цифровая обработка сигналов. — СПб.: Академия, 2013. — 318 с.
2. Гоноровский И. С. Радиотехнические цепи и сигналы. — М.: Радио и связь, 1986. — 512 с.
3. Alan V. Oppenheim, Ronald W. Schaffer. Discrete-Time Signal Processing. 3 rd ed. — Prentice Hall, 2009. — 1120 p.

UDC 621.372

doi:10.15217/issn1684-8853.2016.2.104

Analysis of Linear Systems on the base of Transient Characteristics

Ziatdinov S. I.^a, Dr. Sc., Tech., Professor, kaf53@guap.ru

^aSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., Saint-Petersburg, Russian Federation

Introduction: The transition from continuous linear systems to discrete systems using the readings of impulsive characteristics of similar continuous linear systems leads to mistakes in forming the transient processes. At the same time, the dynamic properties of linear systems are fully determined by their transient characteristics. The synthesis of discrete linear systems on the base of transient characteristics provides the dynamic properties of a discrete linear system which coincide with the dynamic properties of the continuous linear systems. **Purpose:** The goal is to develop a technique for studying linear systems with the use of transient characteristics. **Methods:** The input signal is represented as a sum of step influences shifted one against another. Using the transient characteristic, the output signal of a linear system is found as a sum of the system reactions to each step influence. **Results:** A new form of imposition integral is obtained which allows you, when the input signal and transient characteristic of a linear system are known, to calculate its output signal. The theoretical results are corroborated by examples. **Practical relevance:** The obtained technique allows you to synthesize discrete linear systems whose dynamic properties coincide with the dynamic properties of similar continuous linear systems.

Keywords — Linear System, Transient Characteristic, Imposition Integral.

Reference

1. Vorobyov S. N. *Tsifrovaya obrabotka signalov* [Digital Signal Processing]. Saint-Petersburg, Akademiia Publ., 2013. 318 p. (In Russian).
2. Gonorovsky I. S. *Radiotekhnicheskie tsepi i signaly* [Radio Circuits and Signals]. Moscow, Radio i svyaz' Publ., 1986. 512 p. (In Russian).
3. Alan V. Oppenheim, Ronald W. Schaffer. *Discrete-Time Signal Processing*. 3 rd ed. Prentice Hall, 2009. 1120 p. (In Russian).

АНТОНОВ
Вячеслав
Викторович



Заведующий кафедрой автоматизированных систем управления Уфимского государственного авиационного технического университета. В 1979 году окончил Башкирский государственный университет по специальности «Математика». В 2015 году защитил диссертацию на соискание ученой степени доктора технических наук. Является автором 60 научных публикаций. Область научных интересов — автоматизированные информационные системы, математическое и программное обеспечение компьютерных сетей. Эл. адрес: boss@bashkortostan.ru

БАЛЫШЕВА
Ольга
Леонидовна



Доцент кафедры электроники и оптической связи Санкт-Петербургского государственного университета аэрокосмического приборостроения. В 1992 году окончила Ленинградский институт авиационного приборостроения по специальности «Радиотехника». В 1998 году защитила диссертацию на соискание ученой степени кандидата технических наук. Является автором 40 научных публикаций и одного патента на изобретение. Область научных интересов — акустоэлектроника, радиотехника. Эл. адрес: balysheva@mail.ru

БЕЗРУЧКО
Борис
Петрович



Профессор кафедры динамического моделирования и биомедицинской инженерии Саратовского государственного университета им. Н. Г. Чернышевского. В 1969 году окончил физический факультет Саратовского государственного университета им. Н. Г. Чернышевского по специальности «Радиофизика». В 1995 году защитил диссертацию на соискание ученой степени доктора физико-математических наук. Является автором более 250 научных публикаций. Область научных интересов — стохастическая динамика и хаос, нелинейные колебания и волны, радиофизика, анализ временных рядов. Эл. адрес: bezruchkobp@gmail.com

БАЛОНИН
Николай
Алексеевич



Профессор кафедры вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения. В 1982 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматика и телемеханика». В 2008 году защитил диссертацию на соискание ученой степени доктора технических наук. Является автором более 90 научных публикаций, в том числе трех монографий. Область научных интересов — теория динамических систем, теория идентификации, теория операторов, теория матриц, вычислительные методы, интернет-робототехника, интернет-книги с исполняемыми алгоритмами, научные социальные сети. Эл. адрес: korbendfs@mail.ru

БАСОВ
Олег
Олегович



Докторант Академии федеральной службы охраны Российской Федерации, Орел. В 2004 году окончил Академию службы специальной связи и информации по специальности «Сети связи и системы коммутации». В 2009 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 100 научных публикаций и 25 патентов на изобретения и полезные модели. Область научных интересов — обработка и кодирование речевых и иконических сигналов, проектирование полимодальных инфокоммуникационных систем. Эл. адрес: oobasov@mail.ru

БУРАКОВ
Михаил
Владимирович



Доцент кафедры управления и информатики в технических системах Санкт-Петербургского государственного университета аэрокосмического приборостроения. В 1984 году окончил Ленинградский институт авиационного приборостроения по специальности «Автоматизированные системы управления». В 1994 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором более 150 научных публикаций. Область научных интересов — системы интеллектуального управления, нечеткие регуляторы, нейронные сети, эволюционные алгоритмы. Эл. адрес: bmv@sknt.ru

**ЕГОРОВ
Игорь
Валерьевич**



Аспирант кафедры компьютерных систем и программных технологий Института компьютерных наук и технологий Санкт-Петербургского государственного политехнического университета Петра Великого.

В 2012 году окончил с отличием факультет технической кибернетики Санкт-Петербургского государственного политехнического университета по специальности «Информатика и вычислительная техника».

Является автором трех научных публикаций.

Область научных интересов — надежность вычислительных систем, программная инженерия, проектирование вычислительных систем.

Эл. адрес: iegorov@kspt.icc.spbstu.ru

**КАПЛИН
Александр
Юрьевич**



Заместитель генерального директора, генеральный конструктор ОАО «Радиоавионика», Санкт-Петербург.

В 1985 году окончил Ленинградский механический институт по специальности «Радиоэлектронные устройства».

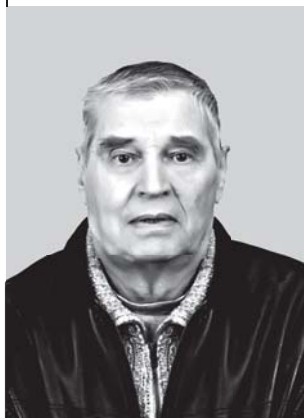
В 1997 году защитил диссертацию на соискание ученой степени кандидата технических наук.

Является автором 50 научных публикаций, двух авторских свидетельств, двух патентов на изобретение и одного патента на полезную модель.

Область научных интересов — системы управления и связи специального назначения, человеко-машинные системы, бортовая радиолокация и радионавигация.

Эл. адрес: a.kaplin@list.ru

**КЛУДЗИН
Виктор
Владимирович**



Профессор кафедры электроники и оптической связи Санкт-Петербургского государственного университета аэрокосмического приборостроения, заслуженный работник высшей школы РФ.

В 1963 году окончил Ленинградский институт авиационного приборостроения по специальности «Радиоэлектроника».

В 1989 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 100 научных публикаций, в том числе трех монографий.

Область научных интересов — акустооптические методы обработки радио- и оптических сигналов, физическая акустика, волновая оптика, радиоэлектроника.

Эл. адрес: weconf@home.ru

**ЗИАТДИНОВ
Сергей
Ильич**



Профессор кафедры информационно-сетевых технологий Санкт-Петербургского государственного университета аэрокосмического приборостроения.

В 1970 году окончил Ленинградский институт авиационного приборостроения по специальности «Электронные устройства».

В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором 140 научных публикаций и 16 патентов на изобретения.

Область научных интересов — аналоговая и цифровая обработка сигналов, автоматические системы управления.

Эл. адрес: kaf53@guap.ru

**КИПЯТКОВА
Ирина
Сергеевна**



Старший научный сотрудник лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского института информатики и автоматизации РАН.

В 2008 году окончила Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Роботы и робототехнические системы».

В 2011 году защитила диссертацию на соискание ученой степени кандидата технических наук.

Является автором 60 научных публикаций.

Область научных интересов — речевые технологии, автоматическое распознавание речи, многомодальные человеко-машинные интерфейсы.

Эл. адрес: kipyatkova@ias.spb.su

**КОЗИНОВ
Игорь
Александрович**



Доцент, докторант кафедры космических радиотехнических систем Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург.

В 1999 году окончил Военно-космическую академию им. А. Ф. Можайского по специальности «Радиоэлектронные системы».

В 2010 году защитил диссертацию на соискание ученой степени кандидата технических наук.

Является автором более 50 научных публикаций.

Область научных интересов — обработка сигналов в радиотехнических и оптико-электронных информационных системах, космические радиотехнические комплексы управления, сбора и передачи информации.

Эл. адрес: garry-spb@yandex.ru

**КОНОВАЛОВ
Александр
Сергеевич**



Профессор кафедры управления и информатики в технических системах Санкт-Петербургского государственного университета аэрокосмического приборостроения, почетный работник высшего профессионального образования РФ.

В 1968 году окончил Ленинградский институт авиационного приборостроения по специальности «Электрооборудование летательных аппаратов».

В 1998 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 120 научных публикаций.

Область научных интересов — автоматическое и интеллектуальное управление.

Эл. адрес: Kononov@c4t.com

**КРАСИЛЬНИКОВ
Николай
Николаевич**



Профессор кафедры информационно-сетевых технологий Санкт-Петербургского государственного университета аэрокосмического приборостроения, заслуженный деятель науки и техники РФ. В 1950 году окончил Ленинградский политехнический институт. В 1963 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 240 научных публикаций, в том числе пяти монографий.

Область научных интересов — цифровая обработка изображений, статистическая теория передачи и восприятия изображений, математическое моделирование процессов обработки информации зрительной системой человека.

Эл. адрес:

NNKrasilnikov@yandex.ru

**КРАСИЛЬНИКОВА
Ольга
Ивановна**



Доцент кафедры информационно-сетевых технологий Санкт-Петербургского государственного университета аэрокосмического приборостроения.

В 1978 году окончила Ленинградский институт авиационного приборостроения по специальности «Радиотехника».

В 1991 году защитила диссертацию на соискание ученой степени кандидата технических наук.

Является автором более 90 научных публикаций и семи учебных пособий.

Область научных интересов — цифровая обработка изображений.

Эл. адрес:

OIKrasilnikova@yandex.ru

**КУЛАКОВ
Сергей
Викторович**



Профессор, заведующий кафедрой электроники и оптической связи Санкт-Петербургского государственного университета аэрокосмического приборостроения, лауреат Государственной премии СССР, заслуженный деятель науки и техники РФ, академик Международной академии наук высшей школы.

В 1953 году окончил Ленинградский институт авиационного приборостроения по специальности «Радиотехника».

В 1980 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором 160 научных публикаций и более 30 авторских свидетельств на изобретения.

Область научных интересов — акустооптика, акустоэлектроника, радиотехника.

Эл. адрес: svk25@mail.ru

**КУЛИКОВ
Геннадий
Григорьевич**



Профессор кафедры автоматизированных систем управления Уфимского государственного авиационного технического университета.

В 1971 году окончил Уфимский авиационный институт по специальности «Автоматизация машиностроения».

В 1989 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором 160 научных публикаций.

Область научных интересов — автоматизированные системы управления и системы управления силовыми установками летательных аппаратов.

Эл. адрес:

gennadyg_98@Yahoo.com

**КУЛЬМИНСКИЙ
Данил
Дмитриевич**



Младший научный сотрудник Саратовского филиала Института радиотехники и электроники им. В. А. Котельникова РАН, аспирант Саратовского государственного университета им. Н. Г. Чернышевского.

В 2014 году окончил Саратовский государственный университет им. Н. Г. Чернышевского по специальности «Методы и устройства обработки сигналов и данных».

Является автором девяти научных публикаций.

Область научных интересов — радиофизика, нелинейная динамика.

Эл. адрес:

kulminskydd@gmail.com

**КУТУЗОВ
Олег
Иванович**



Профессор кафедры автоматизированных систем обработки информации и управления Санкт-Петербургского государственного электротехнического университета «ЛЭТИ», почетный работник высшего профессионального образования РФ.

В 1960 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматика и телемеханика».

В 1996 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 150 научных публикаций.

Область научных интересов — моделирование инфокоммуникационных сетей и их структурообразующих звеньев.

Эл. адрес: kutuzov-oleg@mail.ru

**МАЛЬЦЕВ
Георгий
Николаевич**



Профессор кафедры космических радиотехнических систем Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург, заслуженный деятель науки РФ, действительный член Академии космонавтики им. К. Э. Циолковского.

В 1980 году окончил Военный инженерный Краснознаменный институт им. А. Ф. Можайского. В 1994 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 300 научных публикаций и 27 патентов на изобретения.

Область научных интересов — обработка сигналов в радиотехнических и оптико-электронных информационных системах и др.

Эл. адрес: georgiy_maltsev@mail.ru

**ПОНОМАРЕНКО
Владимир
Иванович**



Доцент, ведущий научный сотрудник Саратовского филиала Института радиотехники и электроники им. В. А. Котельникова РАН.

В 1982 году окончил Саратовский государственный университет им. Н. Г. Чернышевского по специальности «Радиофизика». В 2008 году защитил диссертацию на соискание ученой степени доктора физико-математических наук.

Является автором более 200 научных публикаций.

Область научных интересов — стохастическая динамика и хаос, нелинейные колебания и волны, радиофизика, кодирование и передача информации, анализ временных рядов.

Эл. адрес: ponomarenkovi@gmail.com

**КУЧМИН
Андрей
Юрьевич**



Старший научный сотрудник лаборатории механики управляемых систем Института проблем машиноведения РАН, Санкт-Петербург.

В 2005 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Информатика и управление в технических системах».

В 2007 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 40 научных публикаций и двух патентов на изобретения.

Область научных интересов — математическое моделирование в естественных науках, искусственный интеллект и принятие решений, математические проблемы теории управления и др.

Эл. адрес: radiotelescope@yandex.ru

**МЕЛЕХИН
Виктор
Федорович**



Профессор кафедры компьютерных систем и программных технологий Санкт-Петербургского политехнического университета Петра Великого, заслуженный работник высшей школы РФ.

В 1960 году окончил Ленинградский политехнический институт им. М. И. Калинина по специальности «Автоматика и телемеханика».

В 1984 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором 230 научных публикаций, 52 изобретений, 23 патентов, четырех монографий, шести учебников и 14 учебных пособий.

Область научных интересов — теория и технология проектирования вычислительных систем и устройств.

Эл. адрес: melekhin@kpspt.icc.spbstu.ru

**ПРОСКУРИН
Александр
Викторович**



Аспирант кафедры информатики и вычислительной техники Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнёва, Красноярск.

В 2013 году окончил магистратуру Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнёва по специальности «Информатика и вычислительная техника».

Является автором 20 научных публикаций. Область научных интересов — цифровая обработка изображений, распознавание образов, автоматическое аннотирование изображений.

Эл. адрес: Proskurin.AV.WOF@gmail.com

**ПРОХОРОВ
Михаил
Дмитриевич**



Доцент, заведующий лабораторией моделирования в нелинейной динамике Саратовского филиала Института радиотехники и электроники им. В. А. Котельникова РАН.

В 1992 году окончил Саратовский государственный университет им. Н. Г. Чернышевского по специальности «Радиофизика». В 2008 году защитил диссертацию на соискание ученой степени доктора физико-математических наук.

Является автором более 200 научных публикаций.

Область научных интересов — нелинейные колебания и волны, математическое моделирование, стохастическая динамика и хаос, анализ временных рядов.

Эл. адрес:

mdprokhorov@yandex.ru

**САВЕЛЬЕВ
Антон
Игоревич**



Соискатель, научный сотрудник лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского института информатики и автоматизации РАН.

В 2012 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Роботы и робототехнические системы».

Является автором 13 научных публикаций и одного патента на изобретения и полезные модели.

Область научных интересов — человеко-машинное взаимодействие, робототехника, инфокоммуникационные системы.

Эл. адрес: saveliev@ias.spb.su

**САИТОВ
Игорь
Акрамович**



Доктор технических наук, профессор, начальник факультета Академии Федеральной службы охраны Российской Федерации, Орел.

В 1988 году окончил Орловское высшее военное командное училище связи по специальности «Эксплуатация средств электропроводной связи», в 1998 году — Военную академию связи по специальности «Управление техническим обеспечением систем военной связи».

В 2011 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором 200 научных публикаций и 15 патентов на изобретения и полезные модели.

Область научных интересов — методология синтеза (оптимизации характеристик) и обеспечения эффективного функционирования инфокоммуникационных систем специального назначения.

Эл. адрес: soul1308@yandex.ru

**СЕРГЕЕВ
Михаил
Борисович**



Профессор, заведующий кафедрой вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения, директор НИИ информационно-управляющих систем Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики.

В 1980 году окончил «ЛЭТИ».

В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 100 научных публикаций и 14 патентов на изобретения.

Область научных интересов — теория разрядных вычислений, методы проектирования спецпроцессоров для систем контроля и управления и др.

Эл. адрес: mbse@mail.ru

**СТЕПАНОВ
Михаил
Георгиевич**



Профессор, заместитель генерального конструктора по научной работе ОАО «Радиоавионика», Санкт-Петербург.

В 1971 году окончил Ленинградскую военную инженерную Краснознаменную академию им. А. Ф. Можайского по специальности «Радиотехнические системы».

В 1995 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором 150 научных публикаций, 11 авторских свидетельств и одного патента на изобретение.

Область научных интересов — теория статистического оценивания и обработки информации, космическая радионавигация и радиолокация, системы управления и связи специального назначения.

Эл. адрес: smg099@mail.ru

**ТАТАРНИКОВА
Татьяна
Михайловна**



Профессор кафедры безопасности информационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.

В 1993 году окончила Восточно-Сибирский технологический институт по специальности «Электронно-вычислительные машины, комплексы, системы и сети».

В 2007 году защитила диссертацию на соискание ученой степени доктора технических наук.

Является автором более 100 научных публикаций.

Область научных интересов — инфокоммуникации, взаимодействие неоднородных сетей.

Эл. адрес: tm-tatarn@yandex.ru

ТАУБИН
Феликс
Александрович



Профессор кафедры информационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения. В 1968 году окончил Ленинградский институт авиационного приборостроения по специальности «Радиоэлектронные устройства систем управления». В 1992 году защитил диссертацию на соискание ученой степени доктора технических наук. Является автором более 80 научных публикаций, девяти патентов и двух изобретений. Область научных интересов — цифровые системы связи, методы помехоустойчивого кодирования, широкополосные системы, беспроводные сети. Эл. адрес: ftaubin@yahoo.com

ТРОФИМОВ
Андрей
Николаевич



Доцент кафедры инфокоммуникационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения. В 1976 году окончил Ленинградский институт авиационного приборостроения по специальности «Автоматизированные системы управления». В 1982 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 40 научных публикаций и шести патентов на изобретения. Область научных интересов — теория передачи дискретных сообщений, теория информации, теория кодирования. Эл. адрес: andrei.trofimov@vu.spb.ru

ФАВОРСКАЯ
Маргарита
Николаевна



Профессор, заведующая кафедрой информатики и вычислительной техники Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнёва, Красноярск. Почетный работник высшего профессионального образования РФ. В 1980 году окончила Рыбинский авиационный технологический институт по специальности «Конструирование и производство электронно-вычислительной аппаратуры». В 2011 году защитила диссертацию на соискание ученой степени доктора технических наук. Является автором около 160 научных публикаций. Область научных интересов — распознавание образов, цифровая обработка изображений, кластерный анализ, интеллектуальные технологии обработки данных. Эл. адрес: favorskaya@sibsau.ru

ФАХРУЛЛИНА
Альмира
Раисовна



Старший преподаватель кафедры автоматизированных систем управления Уфимского государственного авиационного технического университета. В 2004 году окончила филиал Уфимского государственного авиационного технического университета в г. Кумертау по специальности «Информационные системы (в экономике)». Является автором 22 научных публикаций. Область научных интересов — математическое и программное обеспечение вычислительных машин, автоматизация процессов управления, методы и технологии обработки данных. Эл. адрес: almirafax@mail.ru

ШИЛИНА
Мария
Анатольевна



Доцент кафедры автоматизированных систем управления Уфимского государственного авиационного технического университета. В 2006 году окончила Уфимский государственный авиационный технический университет по специальности «Прикладная информатика в экономике». В 2012 году защитила диссертацию на соискание ученой степени кандидата технических наук. Является автором 35 научных публикаций. Область научных интересов — автоматизация процессов управления, методы и технологии обработки данных. Эл. адрес: maria.shilina@gmail.com



IX Специализированный форум ПТА. Санкт-Петербург 2016



Передовые Технологии Автоматизации

**1-2
июня**

Санкт-Петербург

КЗ «Конгресс», ул. Шпалерная, 56

В программе форума:

- Автоматизация в эпоху Индустрии 4.0
- Диспетчерское управление инфраструктурными объектами
- Практика автоматизации инженерных систем зданий
- Интеллектуальное здание будущего
- Промышленная автоматизация в проектах и решениях
- Интеллектуальные системы безопасности

Организатор
Экспотромитек

Санкт-Петербург:

Тел.: (812) 448-03-38
E-mail: info@pta-expo.ru

Москва:

Тел.: (495) 234-22-10
E-mail: info@pta-expo.ru

ISSN 1684-8853

