

ISSN 1819-3161

ПРОБЛЕМЫ УПРАВЛЕНИЯ

3/2025

CONTROL  SCIENCES

РЕДАКЦИОННЫЙ СОВЕТ

С. Н. Васильев, академик РАН,
И. А. Каляев, академик РАН,
Н. В. Кузнецов, чл.-корр. РАН,
В. А. Левин, академик РАН,
Н. А. Махутов, чл.-корр. РАН,
А. Ф. Резчиков, чл.-корр. РАН,
Е. А. Федосов, академик РАН

РЕДКОЛЛЕГИЯ

Ф. Т. Алескеров, д-р техн. наук,
В. Н. Афанасьев, д-р техн. наук,
Н. Н. Бахтадзе, д-р техн. наук,
В. М. Вишневский, д-р техн. наук,
О. И. Дранко, д-р техн. наук
А. О. Калашников, д-р техн. наук,
В. В. Клочков, д-р экон. наук,
С. А. Краснова, д-р техн. наук,
О. П. Кузнецов, д-р техн. наук,
А. А. Лазарев, д-р физ.-мат. наук,
В. Г. Лебедев, д-р техн. наук,
В. Е. Лепский, д-р психол. наук,
Н. Е. Максимова, канд. техн. наук
(ответственный секретарь),
А. С. Мандель, д-р техн. наук,
Р. В. Мещеряков, д-р техн. наук,
А. И. Михальский, д-р биол. наук,
Д. А. Новиков, академик РАН
(гл. редактор),
Б. В. Павлов, д-р техн. наук,
Ф. Ф. Пашченко, д-р техн. наук
(зам. гл. редактора),
Л. Б. Рапопорт, д-р физ.-мат. наук,
С. В. Ратнер, д-р экон. наук,
Е. Я. Рубинович, д-р техн. наук,
Л. Ю. Филимонюк, д-р техн. наук
М. В. Хлебников, д-р физ.-мат. наук,
А. Г. Чхартишвили, д-р физ.-мат. наук,
И. Б. Ядыкин, д-р техн. наук

РУКОВОДИТЕЛИ РЕГИОНАЛЬНЫХ РЕДСОВЕТОВ

Владивосток – О. В. Абрамов, д-р техн. наук,
Волгоград – А. А. Воронин, д-р физ.-мат. наук,
Воронеж – С. А. Баркалов, д-р техн. наук,
Курск – С. Г. Емельянов, д-р техн. наук,
Липецк – А. К. Погодаев, д-р техн. наук,
Пермь – В. Ю. Столбов, д-р техн. наук,
Ростов-на-Дону – Г. А. Угольников,
д-р техн. наук,
Самара – М. И. Гераськин, д-р экон. наук,
Саратов – В. А. Кушников, д-р техн. наук,
Тамбов – М. Н. Краснянский, д-р техн. наук,
Уфа – Б. Г. Ильясов, д-р техн. наук,
Челябинск – О. В. Логиновский, д-р техн. наук

ADVISORY BOARD

E. A. Fedosov, RAS¹ Academician,
I. A. Kalyaev, RAS Academician,
N. V. Kuznetsov, RAS Corr. Member,
V. A. Levin, RAS Academician,
N. A. Makhutov, RAS Corr. Member,
A. F. Rezchikov, RAS Corr. Member,
S. N. Vassilyev, RAS Academician

EDITORIAL BOARD

V. N. Afanas'ev, Dr. Sci. (Tech.),
F. T. Aleskerov, Dr. Sci. (Tech.),
N. N. Bakhtadze, Dr. Sci. (Tech.),
A. G. Chkhartishvili, Dr. Sci. (Phys.-Math.),
O. I. Dranko, Dr. Sci. (Tech.),
L. Yu. Filimonuk, Dr. Sci. (Tech.),
A. O. Kalashnikov, Dr. Sci. (Tech.),
V. V. Klochkov, Dr. Sci. (Econ.),
M. V. Khlebnikov, Dr. Sci. (Phys.-Math.),
S. A. Krasnova, Dr. Sci. (Tech.),
O. P. Kuznetsov, Dr. Sci. (Tech.),
A. A. Lazarev, Dr. Sci. (Phys.-Math.),
V. G. Lebedev, Dr. Sci. (Tech.),
V. E. Lepskiy, D. Sc. (Psych.),
A. S. Mandel, Dr. Sci. (Tech.),
N. E. Maximova, Cand. Sci. (Tech),
Executive Editor-in-Chief,
R. V. Meshcheryakov, Dr. Sci. (Tech.),
A. I. Michalski, Dr. Sci. (Biol.),
D. A. Novikov, RAS Academician,
Editor-in-Chief,
F. F. Pashchenko, Dr. Sci. (Tech.),
Deputy Editor-in-Chief,
B. V. Pavlov, Dr. Sci. (Tech.),
L. B. Rapoport, Dr. Sci. (Phys.-Math.),
S. V. Ratner, Dr. Sci. (Econ.),
E. Ya. Rubinovich, Dr. Sci. (Tech.),
A. D. Tsvirkun, Dr. Sci. (Tech.),
V. M. Vishnevsky, Dr. Sci. (Tech.),
I. B. Yadykin, Dr. Sci. (Tech)

LEADERS OF REGIONAL BOARDS

Chelyabinsk – O. V. Loginovskiy, Dr. Sci. (Tech.),
Kursk – S. G. Emelyanov, Dr. Sci. (Tech.),
Lipetsk – A. K. Pogodaev, Dr. Sci. (Tech.),
Perm – V. Yu. Stolbov, Dr. Sci. (Tech.),
Rostov-on-Don – G. A. Ougolnitsky,
Dr. Sci. (Tech.),
Samara – M. I. Geraskin, Dr. Sci. (Econ.),
Saratov – V. A. Kushnikov, Dr. Sci. (Tech.),
Tambov – M. N. Krasnyanskiy, Dr. Sci. (Tech.),
Ufa – B. G. Ilyasov, Dr. Sci. (Tech.),
Vladivostok – O. V. Abramov, Dr. Sci. (Tech.),
Volgograd – A. A. Voronin, Dr. Sci. (Phys.-Math.),
Voronezh – S. A. Barkalov, Dr. Sci. (Tech.)

¹Russian Academy of Sciences.



CONTROL SCIENCES
Научно-технический
журнал

6 номеров в год

ISSN 1819-3161 (Print)

ISSN 2712-8687 (Online)

Издается с 2003 года

УЧРЕДИТЕЛЬ и ИЗДАТЕЛЬ

Федеральное государственное
бюджетное учреждение науки
Институт проблем управления
им. В.А. Трапезникова РАН

Главный редактор
академик РАН
Д.А. Новиков

Заместитель главного редактора
Ф.Ф. Пащенко

Ответственный секретарь
Н.Е. Максимова

Выпускающий редактор
Л.В. Петракова

Адрес редакции
117997, ГСП-7, Москва,
ул. Профсоюзная, д. 65, к. 410

Тел./факс (495) 198-17-20, доб. 1410

E-mail: pu@ipu.ru

Интернет: <http://pu.mtas.ru>
<http://controlsciences.org>

Опубликовано: 16 июля 2025 г.

Свидетельство о регистрации
ПИ № ФС 77-49203 от 30 марта 2012 г.
выдано Министерством Российской
Федерации по делам печати,
телерадиовещания и средств массовых
коммуникаций

Свидетельство о регистрации
Эл № ФС 77-80482 от 17 февраля 2021 г.
выдано Федеральной службой
по надзору в сфере связи,
информационных технологий и
массовых коммуникаций

Журнал входит в RSCI на платформе
Web of Science и Перечень
рецензируемых научных изданий ВАК

Журнал включен в Российский индекс
научного цитирования (РИНЦ).
На сайте Научной электронной
библиотеки (www.elibrary.ru) доступны
полные тексты статей.

© Федеральное государственное
бюджетное учреждение науки
Институт проблем управления
им. В.А. Трапезникова РАН

ПРОБЛЕМЫ УПРАВЛЕНИЯ

3.2025

СОДЕРЖАНИЕ

Обзоры

Буков В. Н., Бронников А. М., Попов А. С., Шурман В. А.

Методы мониторинга технического состояния систем
в целях управления их избыточностью. Ч. 2. Классические
модели 3

Математические проблемы управления

Бабилов О. В., Бабилов В. Г. Оценка фундаментальной

частоты вально-лопастного ряда с применением
преобразования Гильберта и автокорреляции 15

Управление в социально-экономических системах

Гребенков Д. И., Козлова А. А., Лемтюжникова Д. В.,

Новиков Д. А. Модели утомления и отдыха при научении.
Ч. 1. Расширение общей модели итеративного научения 28

Угольников Г. А. Сущность, признаки и принцип

описания организационных систем 38

Информационные технологии в управлении

Ефанов Д. В. Композиции двух равновесных кодов с

ортогональными по всем разрядам комбинациями
для синтеза самопроверяемых дискретных устройств 49

Есин А. А. Математическая модель адаптивного

управления трафиком в подвижных сетях с переменным
качеством сигнала 63



CONTROL SCIENCES
Scientific Technical
Journal

6 issues per year

ISSN 1819-3161 (Print)

ISSN 2712-8687 (Online)

Published since 2003

FOUNDER and PUBLISHER

V.A. Trapeznikov

Institute of Control Sciences
of Russian Academy of Sciences

Editor-in-Chief

D.A. Novikov, RAS Academician

Deputy Editor-in-Chief

F.F. Pashchenko

Executive Editor-in-Chief

N.E. Maximova

Editor

L.V. Petrakova

Editorial address

65 Profsoyuznaya st., office 410,
Moscow 117997, Russia

☎/📠 +7(495) 198-17-20, ext. 1410

✉ pu@ipu.ru

URL: <http://pu.mtas.ru>

<http://controlsciences.org>

Published: July 16, 2025

Registration certificate of

ПИ № ФС 77-49203 of 30 March 2012

issued by the Ministry of Press,

Broadcasting, and Mass Media

of the Russian Federation

Registration certificate of

Эл № ФС 77-80482 of 17 February 2021

issued by the Federal Service

for Supervision of Communications,

Information Technology, and Mass Media

The Journal is indexed in RSCI (Russian

Science Citation Index) on the platform

Web of Science and in the list of peer-

reviewed scientific publications of HAC

On the website of the Scientific electronic

library (www.elibrary.ru) full texts of

articles are available

© V.A. Trapeznikov

Institute of Control Sciences

of Russian Academy of Sciences

CONTROL SCIENCES

3.2025

CONTENTS

Surveys

- Bukov, V. N., Bronnikov, A. M., Popov, A. S.,
and Shurman, V. A.** Technical Condition Monitoring
Methods to Manage the Redundancy of Systems. Part II:
Classical Models 3

Mathematical Problems of Control

- Babikov, O. V. and Babikov, V. G.** Estimating
the Fundamental Frequency of a Propeller Shaft-Blade
Harmonic Series Using the Hilbert Transform
and Autocorrelation 15

Control in Social and Economic Systems

- Grebenkov, D. I., Kozlova, A. A., Lemtyuzhnikova, D. V.,
and Novikov, D. A.** Models of Fatigue and Rest in Learning.
Part I: Extension of the General Iterative Learning Model 28

- Ougolnitsky, G. A.** The Essence, Attributes, and Description
Principle of Organizational Systems 38

Information Technology in Control

- Efanov, D. V.** Compositions of Two Constant-Weight Codes
with Orthogonal Combinations over All Bits for Self-Checking
Discrete Device Design 49

- Esin, A. A.** A Mathematical Model of Adaptive Traffic Control
in Mobile Networks with Variable Signal Quality 63

МЕТОДЫ МОНИТОРИНГА ТЕХНИЧЕСКОГО СОСТОЯНИЯ СИСТЕМ В ЦЕЛЯХ УПРАВЛЕНИЯ ИХ ИЗБЫТОЧНОСТЬЮ.

Ч. 2. Классические модели

В. Н. Буков*, А. М. Бронников**, А. С. Попов***, В. А. Шурман****

*АО «Бортовые аэронавигационные системы», Москва,

**Московский государственный технический университет им. Н. Э. Баумана,

***ВУНЦ ВВС «Военно-воздушная академия им. проф. Н. Е. Жуковского и Ю. А. Гагарина», Воронеж,

****АО «НИИ авиационного оборудования» г. Жуковский Московская обл.

*✉ v_bukov@mail.ru, **✉ bronnikov_a_m@mail.ru, ***✉ saga30@yandex.ru, ****✉ shurman@niiao.ru

Аннотация. Управление избыточностью технической системы предполагает использование процедуры мониторинга – контроля текущего состояния ее составных частей с целью реконфигурирования системы по мере необходимости. В обзоре, состоящем из четырех частей, приводятся современные и вновь развиваемые методы мониторинга в интересах управления избыточностью. Первая часть обзора посвящена общему описанию встроенных средств контроля, схемам голосования и правилам достоверности. Кратко освещены контрольные коды и методы контроля выполнения программ. Во второй части обзора рассматриваются методы диагностирования, основанные на классическом моделировании диагностируемой системы в дискретно-временной и частотной формах. Приводится схема обнаружения отказов Чоу – Вилски, а также определение и процедуры генерирования невязки в задаче диагностирования. Излагаются основные модельные методы диагностирования, использующие ошибки в уравнениях, наблюдатели, согласованность уравнений и избыточные переменные. В завершение кратко освещается проблема робастности методов диагностирования соответствующего типа.

Ключевые слова: мониторинг технического состояния, управление избыточностью, диагностирование, невязка, аналитические модели систем, метод ошибок в уравнениях, наблюдатели, фильтр Бирда, обнаружение отказов по согласованности уравнений, метод избыточных переменных, робастность методов.

ВВЕДЕНИЕ

Модельные (базирующиеся на математических моделях) методы диагностирования¹ основаны на выявлении несоответствия между реальными измерениями входных и выходных сигналов системы и сигналами ее математической модели, выражаемого так называемой невязкой (*residual*). Данные

¹ В данном контексте подразумевается исключительно аналитическое моделирование в «классическом» варианте, при котором центральным звеном является воспроизведение (с той или иной детальностью) законов функционирования диагностируемой системы; подходы с другими типами моделей вынесены в третью и четвертую части обзора.

методы начали активно развиваться с 1970-х гг. В частности, применялись схемы статического или параллельного резервирования [1, 2], которые могут быть получены непосредственно из измерений (аппаратное резервирование) или из аналитических соотношений (аналитическое резервирование). Обзоры этих методов можно найти в работах [3–5].

В этой части обзора освещены методы, наиболее часто применяемые в системах диагностики сложных технических комплексов различного назначения. Их неоспоримым достоинством является прежде всего инженерная прозрачность, т. е. интуитивная ясность протекающих процессов и

причинно-следственных связей для специалистов соответствующей технической области. Эффективность рассмотренных методов зависит от конкретных условий применения, и их выбор определяется функциональным назначением устройства, его структурной организацией и технологическими особенностями изготовления и функционирования, а также требуемыми показателями надежности и достоверности.

1. ОБЩИЕ ПОЛОЖЕНИЯ МЕТОДОВ НА ОСНОВЕ МОДЕЛИРОВАНИЯ

Схема Чоу – Вилски [4], обобщающая большинство методов обнаружения отказного функционирования системы, включающей актуаторы, объект и датчики, показана на рис. 1.

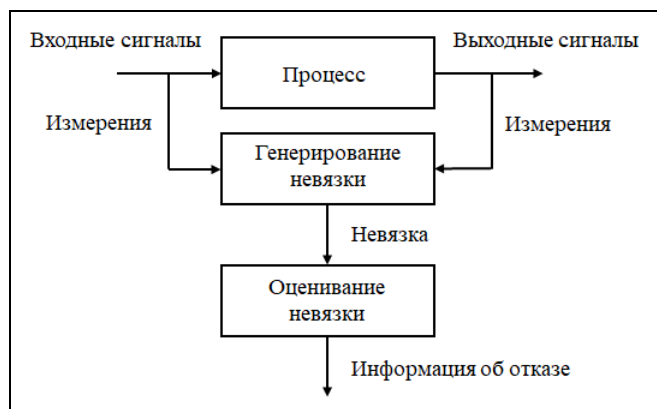


Рис. 1. Схема Чоу – Вилски

Схема обнаружения отказа в реальном времени включает два основных блока:

- Генерирование невязки: путем использования входных и выходных сигналов контролируемого процесса формируется сигнал, называемый невязкой. Эта невязка не должна зависеть от входных и выходных сигналов и при отсутствии отказов должна принимать значение, равное или близкое к нулю.
- Оценивание невязки: проводится анализ значения невязки, на основе которого принимается решение о наличии или отсутствии отказа. Правила принятия решения могут использовать как простой допусковый контроль, так и сложные формы преобразования сигналов, включая методы статистического анализа.

При этом основное внимание исследователей сосредоточено на генерировании невязки, поскольку процедуры ее оценивания в значительной степени универсальны.

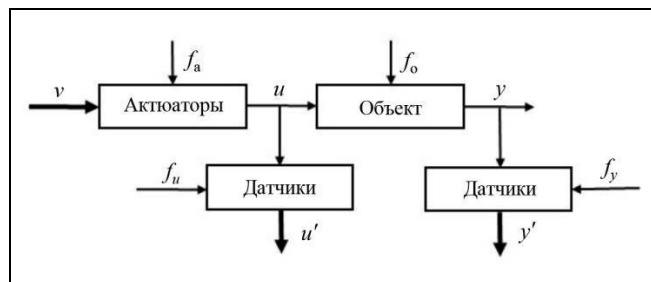


Рис. 2. Топология отказов динамической системы

Топологию отказов, влияющих на работу системы, поясняет рис. 2, где v – внешнее входное воздействие; u – управление, подаваемое на объект; y – выходной сигнал объекта; u' и y' – измеренные (доступные для обработки) сигналы; f_a , f_o , f_u и f_y – формализованное представление отказов, возникающих в актуаторах, объекте и в датчиках входного и выходного сигналов объекта. С той или иной точностью известны только сигналы v , u' и y' .

Хотя большинство систем, подлежащих мониторингу, на самом деле являются нелинейными, методы моделирования и идентификации линейных систем часто используются для того, чтобы избежать сложностей, которые характерны для использования нелинейных моделей. В большинстве случаев это не является существенным ограничением, поскольку мониторинг динамической системы осуществляется в окрестности ожидаемого режима функционирования, а отклонения от него хорошо описываются линейными моделями.

1.1. Описание аналитических моделей

В соответствии с гипотезой линейности динамического процесса его поведение в дискретном времени $t=0, 1, 2, \dots$ в терминах пространства состояний описывается моделью [6–10]

$$x_{t+1} = Ax_t + Bu_t, \quad y_t = Cx_t, \quad (1)$$

где $x_t \in \mathbb{R}^n$ – вектор состояния системы; $u_t \in \mathbb{R}^r$ – вектор истинных входных сигналов; $y_t \in \mathbb{R}^m$ – вектор истинных выходных сигналов; A , B и C – числовые матрицы коэффициентов системы.

Влияние отказов на объект описывается моделью (вектор отказов f_o представлен двумя векторами $f_{ab,t}$ и $f_{c,t}$, связанными с номером такта t)

$$x_{t+1} = Ax_t + Bu_t + f_{ab,t}, \quad y_t = Cx_t + f_{c,t}, \quad (2)$$



где $f_{ab,t} = \Delta A x_t + \Delta B u_t$, $f_{c,t} = \Delta C x_t$; ΔA , ΔB , ΔC – изменения матриц коэффициентов вследствие возникших отказов. Иногда отказы трактуют иначе: $f_{ab,t} = A \Delta x_t + B \Delta u_t$, $f_{c,t} = \Delta y_t$, где Δx_t – постороннее воздействие на внутренние переменные; Δu_t – ошибки (отказы) актюаторов; Δy_t – ошибки (отказы) датчиков.

Поскольку описания пространства состояний обеспечивают общие и математически строгие инструменты для моделирования системы и робастной генерации невязок как для детерминированного случая (измерения без помех), так и для стохастического (измерения, подверженные шуму), матрицы системы A , B и C в канонической форме для моделей (1) и (2) могут быть получены с помощью процедур многомерной идентификации.

Измерения входных и выходных сигналов подвергаются влиянию отказов согласно формулам

$$u'_t = u_t + f_{u,t}, \quad y'_t = y_t + f_{y,t},$$

где u'_t и y'_t – измеренные значения входных и выходных сигналов.

Обычно для моделирования внезапных и развивающихся отказов используются ступенчатые и темповые (градиентные) сигналы, представляющие собой смещение (одноразовое изменение на одном такте) и дрейф (монотонные изменения на последовательности тактов) соответственно.

Помехи измерений, влияющие на выходные сигналы датчиков, обычно описываются некоррелированными гауссовскими процессами.

Рассматриваемая общая задача диагностирования отказов (англ. *Fault Detection*, FD) на основе модели может быть решена только при знании измеренных последовательностей u'_t и y'_t . Кроме того, широко используются априорные знания о характеристиках получаемых сигналов u'_t и y'_t . В качестве примера можно привести спектр, динамический диапазон сигнала и его вариации. Однако необходимость априорной информации о контролируемых сигналах и зависимость характеристик сигнала от неизвестных условий работы диагностируемой системы являются основными недостатками методов такого класса.

Наряду с моделями (1) и (2) для описания функционирования диагностируемой системы в частотной области могут использоваться модели вида

$$y_t = W_y^u(z) u_t + \Delta W_y^u(z) u_t, \quad (3)$$

где z – оператор сдвига на один такт вперед²; $W_y^u(z)$ – передаточная (рациональная полиномиальная) матрица от входного воздействия u_t к выходу y_t системы; $\Delta W_y^u(z)$ – изменение передаточной матрицы $W_y^u(z)$, вызванное отказами.

Обычно модель (3) применяется в тех случаях, когда для воздействующих отказов, а также возмущений известны частотные характеристики, что позволяет использовать информацию в частотных спектрах в качестве критериев для обнаружения отказов.

1.2. Формирование невязок

Общая структура генератора невязки [11, 12] показана на рис. 3. Здесь z_t – вспомогательный сигнал, который, как правило, содержательно значим для разработчика (например, вектор состояний системы или ее частей), r_t – сигнал невязки, $W_1(u_t, y_t)$ и $W_2(z_t, y_t)$ – основные блоки генератора.



Рис. 3. Общая структура генератора невязки

Независимо от того, какой метод применяется, процесс генерирования невязки представляет собой не что иное, как линейное отображение, входы которого состоят из входов и выходов контролируемого процесса. В безотказном состоянии выполняется условие $r_t = 0$.

Простейший генератор невязки получается, когда передаточная матрица W_1 идентична модели (3) объекта: $W_1(u_t, y_t) = W_y^u(z)$, т. е. это описание реального процесса, полученное в результате процедуры идентификации системы (например, авторегрессионная экзогенная модель [13]).

Наиболее простой и широко используемый способ обнаружения отказов заключается в непосредственном сравнении невязки r_t или ее некото-

² Такое толкование предполагает игнорирование начальных условий модели. При альтернативном толковании оператора как z -преобразования учитываются начальные условия.

рой функции $J(r_i)$ с фиксированным порогом ε или пороговой функцией ε_i следующим образом:

$$J(r_i) = \begin{cases} \leq \varepsilon_i & \text{при } f_i = 0, \\ > \varepsilon_i & \text{при } f_i \neq 0, \end{cases} \quad (4)$$

где f_i – обобщенный вектор отказов. Таким образом, если невязка превышает установленное пороговое значение, то объявляется о наличии отказа.

Такой тест особенно результативен с фиксированными порогами ε , если объект работает преимущественно в устойчивом состоянии. Он реагирует либо после большого внезапного, либо после длительно постепенно нарастающего отказа.

В большинстве практических случаев параметры процесса неизвестны вообще или известны недостаточно точно. Тогда их можно определить с помощью методов оценки параметров, измеряя входные и выходные сигналы u_i и y_i , если известна основная структура модели.

2. МЕТОД ОШИБОК В УРАВНЕНИЯХ

Дискретно-временная модель SISO-объекта (англ. *Single Input Single Output* – один вход, один выход) порядка n записывается в векторной форме

$$y_i = \Psi^T \Theta,$$

где $\Theta = [a_1 \ \dots \ a_n \ b_1 \ \dots \ b_n]^T$ – вектор параметров объекта; $\Psi = [y_{i-1} \ \dots \ u_{i-n}]^T$ – вектор дискретно-временных данных на n последовательных тактах измерений.

Вводится сигнал рассогласования

$$e_i = y_i - \Psi^T \hat{\Theta},$$

где $\hat{\Theta}$ – оценка параметров. По методу наименьших квадратов, связанному с выполнением равенств

$$J(\hat{\Theta}) = \sum_i e_i^2 = e^T e, \quad \frac{dJ(\hat{\Theta})}{d\hat{\Theta}} = 0,$$

оценка параметров определяется формулой

$$\hat{\Theta}_i = (\Psi^T \Psi)^{-1} \Psi^T y_i. \quad (5)$$

Как описано в работах [14–16], оценка (5) может быть также вычислена с помощью рекуррентного³ алгоритма наименьших квадратов в режиме

реального времени $t = 0, 1, 2, \dots$ относительно оценок в момент t :

$$\hat{\Theta}_{t+1} = \hat{\Theta}_t + \gamma_t (y_{t+1} - \Psi_{t+1}^T \hat{\Theta}_t),$$

где

$$\gamma_t = (\Psi_{t+1}^T P_t \Psi_{t+1} + 1)^{-1} P_t \Psi_{t+1}, \quad P_{t+1} = (I_{2n} - \gamma_t \Psi_{t+1}^T) P_t,$$

I_{2n} – единичная матрица размера $2n \times 2n$. В качестве начальных условий P_0 для матрицы P_t принимается либо единичная матрица, либо предполагаемая ковариационная матрица начальных ошибок оценки $\hat{\Theta}_0$.

Для улучшения сходимости можно применять методы фильтрации. В частности, когда на измерения влияют помехи, для оценки параметров может быть использован фильтр Калмана [17, 18].

Собственно диагностирование осуществляется по различию вектора полученных оценок $\hat{\Theta}_i$ параметров и его «опорных» значений, соответствующих либо отсутствию, либо наличию определенных отказов в системе. Для этого могут привлекаться подходы, описанные в § 5 первой части [19] и в третьей части данного обзора.

3. ПОДХОДЫ НА ОСНОВЕ НАБЛЮДАТЕЛЕЙ

Основная идея методов на основе наблюдателей или фильтров заключается в оценке выходов системы по результатам измерений с помощью наблюдателей Люенбергера в детерминированном случае или фильтров Калмана в случае наличия случайных помех. При этом ошибка оценки выхода (или ее взвешенное значение) используется в качестве невязки.

Следует отметить, что при использовании наблюдателя для целей диагностирования необходимо оценивать выходы, а оценка вектора состояния обычно не требуется [20]. Кроме того, преимуществом использования наблюдателя является гибкость в выборе его коэффициентов усиления, что приводит к богатому разнообразию схем диагностирования [21–30].

В линейной динамической модели (1) при точно известных матрицах A , B и C для восстановления переменных используется наблюдатель

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + He_t, \quad e_t = y_t - C\hat{x}_t, \quad (6)$$

где $\hat{x}_t \in \mathcal{R}^n$ – оценка вектора состояния; $e_t \in \mathcal{R}^m$ – ошибка выхода. В то же время ошибка состояния $e_t^x \in \mathcal{R}^n$ удовлетворяет уравнениям

$$e_t^x = x_t - \hat{x}_t, \quad e_{t+1}^x = (A - HC)e_t^x$$

³ В программировании используется эквивалентный термин «рекурсивный».



и исчезает асимптотически, если матрица $A - HC$ устойчивая.

В случае наличия возмущений и отказов справедливы уравнения

$$x_{t+1} = Ax_t + Bu_t + Qv_t + L_1 f_t, \quad y_t = Cx_t + R w_t + L_2 f_t,$$

где v_t – неизмеряемый вектор возмущений на входе; w_t – неизмеряемый вектор возмущений на выходе; f_t – приведенный вектор отказов, аддитивно влияющих на функционирование объекта; Q и R – матрицы влияния возмущений; L_1 и L_2 – матрицы влияния отказов.

Тогда для оценки состояния при условиях $v_t = 0$ и $w_t = 0$ справедливы уравнения

$$e_{t+1}^x = (A - HC)e_t^x + L_1 f_t - HL_2 f_t, \\ e_t = Ce_t^x + L_2 f_t,$$

которые описывают влияние обобщенных отказов f_t на ошибки e_t^x и e_t , используемые в качестве невязок. Выбор матрицы H влияет на свойства невязок. При этом ограничивающими условиями выбора H являются устойчивость и чувствительность к возмущениям v_t и w_t . Если на сигналы влияют помехи, то вместо классических наблюдений необходимо использовать фильтр Калмана [17].

Если отказы проявляются в виде изменений ΔA или ΔB параметров, то поведение процесса приобретает вид

$$x_{t+1} = (A + \Delta A)x_t + (B + \Delta B)u_t, \quad y_t = Cx_t,$$

в то время как ошибки оценок представлены в виде

$$e_{t+1}^x = (A - HC)e_t^x + \Delta A x_t + \Delta B u_t, \quad e_t = Ce_t^x.$$

В этом случае изменения ΔA и ΔB являются мультипликативными ошибками, и изменения ошибок зависят как от изменений параметров, так и от изменений входных переменных и переменных состояния. Следовательно, влияние изменений параметров на ошибки не так однозначно, как в случае аддитивных дефектов f_t . Конкретные варианты решения можно найти в работах [14, 25].

Для МИМО-процессов (англ. *Multiple Input Multiple Output* – много входов, много выходов) применяется следующая декомпозиция: выделяется либо один наблюдатель, возбуждаемый одним выходом [26], либо банк наблюдателей, возбуждаемых всеми выходами с проверкой гипотез [27], либо банк наблюдателей, возбуждаемых отдельными выходами [26], либо всеми выходами, кроме

одного [28, 29]. Это позволяет обнаружить отказы различных комбинаций датчиков.

В случае МИМО обратная связь H наблюдателя состояния в уравнении (6) выбирается таким образом, чтобы конкретные сигналы отказов $L_1 f_t$ изменялись в определенном направлении, а сигналы отказов $L_2 f_t$ – в определенной плоскости [29, 31].

При наличии направленных векторов невязок задача выделения отказа сводится к определению того, к какому из известных направлений сигнатуры отказов ближе всего лежит вектор невязки. Первоначальная форма «фильтра обнаружения отказов» была предложена Р. Бирдом [31] и Х. Джонсом [32] для генерации направленных векторов невязок.

В дальнейшем было предложено множество более простых методов, включая методы получения «робастного фильтра обнаружения отказов» [33]. Такой фильтр представляет собой класс наблюдателей Люенбергера со специально разработанной матрицей усиления обратной связи. Другой возможностью является использование выходных наблюдателей или так называемых обобщенных наблюдателей, например, наблюдателей с неизвестным входом с реконструкцией выходных сигналов, если оценка переменной состояния $\hat{x}_t$ не представляет основного интереса.

4. ФИЛЬТР БИРДА

Возможности и ограничения подходов, использующих направления сигнатуры отказа, проиллюстрируем фильтром Бирда [31].

Рассматривается линейная динамическая система с дискретным временем $t = 1, 2, \dots$, поведение которой описывается уравнением

$$x_{t+1} = Ax_t + Bu_t + B[0 \quad \dots \quad f_i \quad \dots \quad 0]^T, \quad (7) \\ x_{t=0} = x_0,$$

где $x_t \in \mathbb{R}^n$ – вектор состояния; $u_t \in \mathbb{R}^l$ – вектор управления; $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times l}$ – матрицы постоянных коэффициентов; дополнительное слагаемое $B[0 \quad \dots \quad f_i \quad \dots \quad 0]^T$ соответствует неизвестному воздействию f_i на i -м входе системы, $i = \overline{1, l}$, $l \leq n$. Предполагается, что все рассматриваемые отказы системы сводятся к совокупности неизвестных воздействий на входах.

Использование фильтра Бирда позволяет как обнаружить наличие m отказов указанного типа, так и указать конкретный столбец b_i матрицы B , которому соответствует постороннее воздействие.

Фильтр предполагает введение вспомогательной динамической системы

$$w_{t+1} = Aw_t + Bu_t + K(x_t - w_t), \quad w_{t=0} = w_0, \quad (8)$$

которая повторяет структуру контролируемой системы (7) с добавлением слагаемого $K(x_t - w_t)$. Матрица коэффициентов K здесь выбирается специальным образом.

Вычитая уравнение (8) из уравнения (7), получаем

$$x_{t+1} - w_{t+1} = A(x_t - w_t) - K(x_t - w_t) + B(u_t - u_t) + B \begin{bmatrix} 0 \\ f_i \\ 0 \end{bmatrix}. \quad (9)$$

Вводя обозначение невязки $r_t = x_t - w_t$, запишем уравнение (9) в виде

$$r_{t+1} = (A - K)r_t + B \begin{bmatrix} 0 \\ f_i \\ 0 \end{bmatrix}. \quad (10)$$

Заметим, что невязка не зависит от входного сигнала u_t и состояния системы x_t , а определяется лишь наличием постороннего сигнала f_i и свойствами матрицы $(A - K)$.

Пусть матрица коэффициентов K выбрана таким образом, чтобы выполнялось равенство

$$A - K = \alpha I_n, \quad (11)$$

где I_n – единичная матрица размера $n \times n$; α – действительное число с нормой $|\alpha| < 1$. Тогда формула (10) принимает вид

$$r_{t+1} = \alpha r_t + B \begin{bmatrix} 0 \\ f_i \\ 0 \end{bmatrix}, \quad r_{t=0} = r_0, \quad (12)$$

и фильтр является устойчивым. Это обстоятельство, а также диагональность матрицы (11) приводят к тому, что с течением времени невязка r_t приближается к значению вектора

$$r_{\lim} = \frac{1}{1-\alpha} [b_{1i} \quad \dots \quad b_{ii} \quad \dots \quad b_{ni}]^T f_i.$$

Матрица B уравнения (7) содержит m столбцов, каждый из которых задает в n -мерном пространстве свое направление⁴. Вопрос, таким образом, сводится к определению тех столбцов матрицы B , на которые в векторном представлении проектируется вектор (12). Оцениваться могут одновременно все l воздействий (отказов) f_i .

⁴ Предполагается, что столбцы линейно независимы.

Этим возможности фильтра Бирда не исчерпаны. Путем аналогичных выкладок можно получить соотношения для постолбцового определения изменения элементов матрицы A в уравнении (7). В этом случае следует рассмотреть уравнение

$$x_{t+1} = Ax_t + Bu_t + A[0 \quad \dots \quad \mu_i \quad \dots \quad 0]^T. \quad (13)$$

При выполнении соответствующих условий с течением времени невязка будет стремиться к значению

$$r_{\lim} = \frac{1}{1-\alpha} [a_{1i} \quad \dots \quad a_{ii} \quad \dots \quad a_{ni}]^T \mu_i.$$

В дальнейшем все аналогично изложенному выше. Одновременно могут оцениваться все n воздействий (отказов) μ_i .

Основные свойства фильтра Бирда заключаются в следующем:

- обнаруживаются постоянные отказы указанного типа,
- для получения оценки отказа требуется время для завершения переходных процессов в фильтре,
- при доступности для измерения только части вектора состояния $y_t = Cx_t$, где $y_t \in \mathbb{R}^m$, $C \in \mathbb{R}^{m \times n}$, возможно оценивание только m отказов, описываемых m столбцами по схемам (6) и (12) из множества всех столбцов матриц A и B .

Для дальнейшего изложения воспользуемся формулами, приведенными в § 5 первой части обзора [19].

Пусть синдром представляет собой m -мерный вектор $\Delta = [\Delta_1 \quad \dots \quad \Delta_m]^T$, а возможные отказы объекта формализуются n -мерным вектором $F = [f_1 \quad \dots \quad f_n]^T$. Тогда

$$\Delta = \Phi(Y, Z, F) = 0. \quad (14)$$

Для проявления реакции синдрома Δ на единичный отказ $f_i \neq 0$ достаточно выполнения условия ненулевой чувствительности

$$\exists j: \frac{\partial \Phi_j}{\partial f_i} \neq 0. \quad (15)$$

В случае же множественного отказа, когда одновременно присутствуют несколько отказов f_i , требуется выполнение дополнительного условия невозможности взаимной компенсации (14) эффектов этих отказов

$$\begin{bmatrix} \partial \Phi_1 / \partial f_1 & \dots & \partial \Phi_1 / \partial f_n \\ \vdots & \ddots & \vdots \\ \partial \Phi_m / \partial f_1 & \dots & \partial \Phi_m / \partial f_n \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} \neq 0. \quad (16)$$

Фильтр Бирда является частным случаем подхода с алгебраическими инвариантами, где избыточность создается искусственно введением дополнительной системы (8). Условие чувствительности (15) выполняется на каждом такте всегда, поскольку в качестве синдрома (14) выступает невязка r_{i+1} , а с учетом уравнения (12) $\partial r_{i+1} / \partial f_i = B$ и с учетом уравнения (13) $\partial r_{i+1} / \partial \mu_i = A$. Условие же (16) выливается в оговоренное выше свойство ограниченности числа обнаруживаемых отказов, поскольку оно предписывает линейную независимость столбцов $[\partial r_{1,i+1} / \partial f_k \cdots \partial r_{m,i+1} / \partial f_k]^T$, обеспечивающую невырожденность соответствующей матрицы и, как необходимое условие, равенству числа невязок $r_{j,i+1}$ и обнаруживаемых отказов f_k .

5. ОБНАРУЖЕНИЕ ОТКАЗОВ ПО СОГЛАСОВАННОСТИ УРАВНЕНИЙ

Основная идея подхода на основе согласованности уравнений (англ. *fault detection via parity equations*) состоит в проверке соответствия различных (как правило, входных и выходных) измерений, получаемых на контролируемом объекте и схеме контроля [34].

Как и для наблюдателей, параметры модели и структура наблюдаемого процесса должны быть известны априори.

Принцип обнаружения отказа заключается в следующем. Для SISO-объекта с известной передаточной функцией $W(z) = b(z) / a(z)$ параллельно используется модель $W_m(z) = b_m(z) / a_m(z)$, где $a(z)$, $b(z)$, $a_m(z)$ и $b_m(z)$ – известные полиномы оператора z . По решению разработчика реализуется какая-либо из схем, приведенных на рис. 4.

Схема, приведенная на рис. 4, а, соответствует невязке выходов, на рис. 4, б – невязке входов, а на

рис. 4, в – промежуточному варианту. В этих схемах при совпадении динамических свойств (параметров полиномов) объекта и его модели сигнал невязки r_i равен нулю. Наличие же отказов, приводящих к изменению значений параметров передаточной функции объекта, порождает возникновение сигнала невязки, отличного от нуля.

Представленные схемы обладают различной чувствительностью к различным отказам. Кроме того, для достижения приемлемой чувствительности или грубости они могут быть дополнены различными фильтрами [34, 35]. Здесь по умолчанию измеренные сигналы равны их истинным значениям: $u'_i = u_i$ и $y'_i = y_i$.

В работе [36] показано, что схемы согласованности обеспечивают меньшую гибкость проектирования по сравнению с методами, основанными на использовании наблюдателей и не имеющими каких-либо ограничений.

Одним из недостатков данного метода является то, что для SISO-объекта формируется только один признак (невязка), что не позволяет уточнять характер и место отказа: есть одно уравнение, и ненулевое значение одной невязки свидетельствует только о рассогласовании этого уравнения. Ничего больше из этого извлечь невозможно. Для преодоления указанного недостатка предложены различные решения [33, 37].

6. МЕТОД ИЗБЫТОЧНЫХ ПЕРЕМЕННЫХ

Данный метод предложен Л. А. Мироновским [38] и относится к наиболее результативным. Основой метода является введение динамической системы для переменных $z \in \mathbb{R}^k$, дополняющей модель контролируемого объекта

$$x_{t+1} = Ax_t + Bu_t, \quad (17)$$

где $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^l$, а A и B – постоянные матрицы.

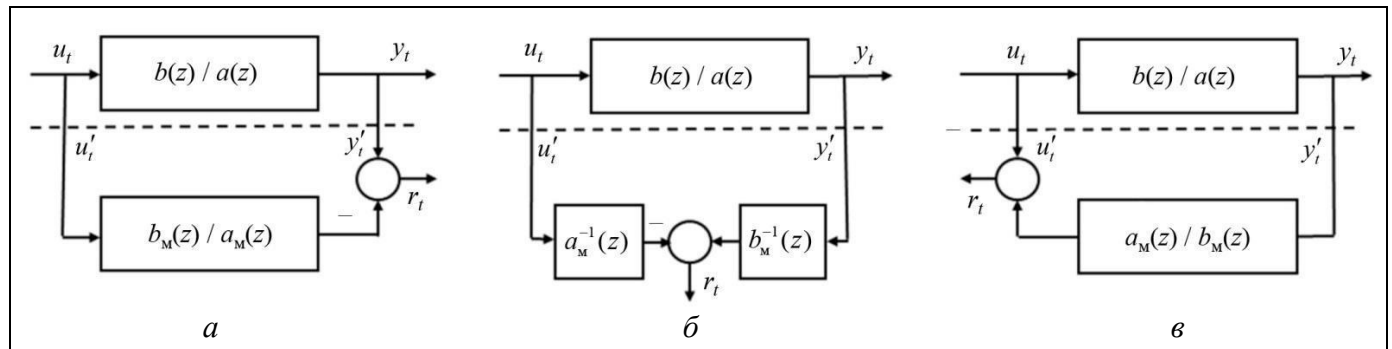


Рис. 4. Типовые схемы согласованности уравнений для обнаружения отказов: а – по выходам, б – по промежуточному сигналу, в – по входам

Избыточные переменные вводятся с удовлетворением двух условий

$$Mz_t = 0, \quad Tz_t = x_t, \quad (18)$$

где $M \in \mathbb{R}^{k \times (n+k)}$ и $T \in \mathbb{R}^{n \times (n+k)}$ – известные (задаваемые) постоянные матрицы⁵. Подставляя выражение (17) в условия (18) и разрешая относительно z_{t+1} , получаем уравнение $(n+k)$ -го порядка

$$z_{t+1} = \begin{bmatrix} T \\ M \end{bmatrix}^{-1} \begin{bmatrix} AT \\ 0 \end{bmatrix} Tz_t + \begin{bmatrix} T \\ M \end{bmatrix}^{-1} \begin{bmatrix} B \\ 0 \end{bmatrix} u_t,$$

описывающее поведение избыточной системы. Эта система, если устойчива, эквивалентна исходной системе (17) в смысле выхода x_t , но одновременно позволяет контролировать ее функционирование по k -мерному вектору невязки

$$r_t = Mz_t.$$

Известны различные уточнения формулировки и решения метода избыточных переменных [11, 39], распространяющие его на модели с передаточными функциями и нелинейные модели, а также позволяющие минимизировать избыточность системы.

Поскольку этот метод, как и фильтр Бирда, является частным случаем метода с алгебраическими инвариантами, то на него распространяются условия чувствительности (15) и (16).

Если возможные отказы объекта контроля формализуются уравнением

$$x_{t+1} = Ax_t + Bu_t + F,$$

где $F = [f_1 \ \dots \ f_n]^T$ – вектор неизмеряемых воздействий, представляющих неучтенные изменения (отказы) в объекте, то в соответствии с условиями (18) получаем на первом такте для каждого из формализованных отказов f_i :

$$\frac{\partial r_{t+1}}{\partial f_i} = \begin{bmatrix} \partial r_{1,t+1} / \partial f_i \\ \vdots \\ \partial r_{k,t+1} / \partial f_i \end{bmatrix} = M \begin{bmatrix} T \\ M \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1_i \\ 0 \\ 0_{k \times 1} \end{bmatrix}_{(n+k) \times 1}, \quad (19)$$

где 1_i – единица, стоящая в i -й строке последнего сомножителя-столбца. А на каждом последующем такте $t+q$, $q=2, \bar{N}$, с учетом формулы (17) получаем

$$\frac{\partial r_{t+q}}{\partial f_i} = \begin{bmatrix} \partial r_{1,t+q} / \partial f_i \\ \vdots \\ \partial r_{k,t+q} / \partial f_i \end{bmatrix} = \sum_{s=0}^{N-1} A^s M \begin{bmatrix} T \\ M \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1_i \\ 0 \\ 0_{k \times 1} \end{bmatrix}_{(n+k) \times 1}, \quad (20)$$

где A^s – s -ая степень матрицы A из формулы (17).

При этом ненулевые значения компонент (19) и (20) при наличии отказов не являются очевидными и зависят от выбора матриц T и M , что требует от разработчика определенного искусства. Так, читатель может убедиться, что в простейшем случае диагональной или кодиагональной матрицы $\begin{bmatrix} T^T & M^T \end{bmatrix}^T$ чувствительность всех компонент невязки r_{t+1} к любому отказу f_i имеет нулевое значение.

7. РОБАСТНОСТЬ МОДЕЛЬНЫХ МЕТОДОВ

Особого внимания заслуживает проблема робастности алгоритмов диагностирования к различным неопределенностям, включая ошибки моделирования и действующие шумы измерителей.

Как правило, в теоретических исследованиях и практических приложениях все неопределенности обобщаются в аддитивные возмущения, действующие на систему (объект, датчики, актуаторы). Хотя вектор возмущений неизвестен, полагается, что оценка его значения может быть получена с помощью процедур идентификации.

Целью подходов к устранению возмущений в системе диагностирования является полное исключение влияния возмущений на невязку, что в общем случае может оказаться невозможным. Существует компромисс между чувствительностью к ошибкам и устойчивостью к неопределенности моделирования, поэтому робастное формирование невязок можно рассматривать как многоцелевую оптимизационную задачу [4, 40]. Она состоит из максимизации влияния дефектов и минимизации влияния неопределенности.

Один из активных способов достижения робастности обнаружения отказов [1] основан на приближенном представлении в модели (1) ошибок моделирования:

$$\Delta W_y^u(z)u_t \approx W_y^d(z)d_t,$$

⁵ Подразумевается, что строки матриц M и T линейно независимы.



где d_t – неизвестный вектор, а $W_y^d(z)$ – оценочная передаточная функция. Если использовать эту приближенную структуру для проектирования генераторов невязок, отключающих помехи, то можно получить робастные алгоритмы обнаружения отказов.

Другой подход, называемый пассивной робастностью, предполагает использование генератора невязок с адаптивным порогом. В качестве простого примера можно привести метод робастного обнаружения отказов с пороговым адаптером или селектором [41]. Этот метод не предполагает никаких усилий для разработки робастной невязки.

В этом случае неопределенность генерации невязок представляется в виде

$$r_t = H(z)\Delta W_y^u(z)u_t.$$

В предположении малости ошибок моделирования

$$\|\Delta W_y^u(z)\| \leq \delta$$

адаптивный порог может быть сформирован линейной системой

$$\varepsilon_t = \delta H(z)u_t.$$

В этом случае порог ε_t уже не фиксирован, а зависит от входного сигнала u_t , являясь, таким образом, адаптивным к рабочему режиму системы. Сбой обнаруживается, если

$$\|r_t\| > \|\varepsilon_t\|.$$

Более детальный подход к робастной генерации невязки предполагает дискретное описание объекта в терминах передаточных матриц

$$y_t = (W_y^u(z) + \Delta W_y^u(z))u_t + W_y^d(z)d_t + W_y^f(z)f_t, \quad (21)$$

где u_t и y_t – вход и выход объекта; f_t – вектор отказов, подлежащих определению; d_t – вектор возмущений; $\Delta W_y^u(z)$ – ошибка представления передаточной матрицы $W_y^u(z)$; $W_y^d(z)$ – влияние возмущения моделирования; матрицы $\Delta W_y^u(z)$ и $W_y^d(z)$ вместе описывают неопределенности моделирования.

Генератор невязок r_t , приведенный на схеме рис. 4, а, описывается уравнением

$$r_t = H_y(z)(y_t - y_t^m), \quad (22)$$

где y_t^m – выход модели (1) без отказов, ошибок и возмущений; $H_y(z)$ – передаточная матрица обра-

ботки невязки. После подстановки выражения (21) в уравнение (22) получаем

$$r_t = H_y(z)W_y^f(z)f_t + H_y(z)W_y^d(z)d_t + H_y(z)\Delta W_y^u(z)u_t. \quad (23)$$

Выделение первого слагаемого в формуле (23) на фоне двух других слагаемых является весьма сложной задачей. Поэтому обычно робастная генерация невязки сводится к выполнению условия

$$H_y(z)W_y^d(z) = 0, \quad (24)$$

для чего применяются различные подходы (с наблюдателями, оптимизационные, с заданными структурами, схемы идентификации и пр.). Часто полное решение такой задачи недостижимо: утрачивается чувствительность к отказам. Однако известны компромиссные решения.

Развязка от возмущений может быть также достигнута с помощью методов проектирования в частотной области. В качестве примера можно привести решение задачи робастного обнаружения отказов с помощью стандартной формулировки фильтрации H_∞ .

Так, в случае невыполнения условия (24) можно получить приближенную оценку, например, в виде индекса эффективности [42]:

$$J_d = \frac{\|H_y(z)W_y^d(z)\|}{\|H_y(z)W_y^f(z)\|} \text{ или } J = \left\| \frac{\partial r}{\partial \varepsilon} \right\| / \left\| \frac{\partial r}{\partial f} \right\|.$$

Более элегантные и продвинутые методы оптимизации H_∞ основаны на использовании алгебраического уравнения Риккати [43]. Часто для формирования невязки используется несколько модифицированный H_∞ -фильтр, т. е. целью проектирования является минимизация влияния возмущений и ошибок моделирования на ошибку оценивания и, как следствие, на невязку. Однако робастное формирование невязки отличается от робастного оценивания тем, что требуется не только ослабление возмущений. Невязка должна оставаться чувствительной к отказам при минимизации влияния возмущений.

ЗАКЛЮЧЕНИЕ

Из второй части обзора следует, что методы диагностирования на основе классического моделирования объекта контроля по-прежнему привлекают значительное количество исследователей. Активно используются методы диагностирования на основе идентификационного подхода, наблюдателей, согласованности уравнений. Широко из-

вестность получил фильтр Бирда, являющийся частным случаем подхода с алгебраическими инвариантами, где избыточность создается искусственно введением дополнительной системы. Одним из результативных подходов к диагностике является метод избыточных переменных. Особого внимания заслуживает проблема робастности алгоритмов диагностирования к различным неопределенностям, включая ошибки моделирования и действующие шумы измерителей.

В третьей части обзора будут изложены методы диагностирования, основанные на нейронных сетях, нечетких и структурных моделях и моделях в виде множеств. Четвертая часть будет посвящена новым подходам к диагностированию и объединению различных моделей и методов.

ЛИТЕРАТУРА

1. Potter, I.E., Suman, M.C. Thresholdless Redundancy Management with Array of Skewed Instruments: Technical Report // AGARDOGRAPH. Integrity in Electronic Flight Control Systems. – 1977. – No. 224. – P. 1511–1525.
2. Авторское свидетельство SU 475604 A1. Устройство для определения границы области работоспособности непрерывной n -мерной технической системы : заявл. 01.02.1973 : опубл. 30.06.1975, бюл. № 24 / Иванов Н.В., Краснов И.А., Мозгалеvский А.В., Юдин Ю.В. [Avtorskoe svidetel'stvo SU 475604 A1. Ustrojstvo dlya opredeleniya granicy oblasti rabotosposobnosti nepreryvnoj n -mernoj tekhnicheskoy sistemy : yayavl. 01.02.1973 : opubl. 30.06.1975, byul. № 24 / Ivanov, N.V., Krasnov, I.A., Mozgalevskij, A.V., Yudin, Yu.V. (In Russian)]
3. Ray, A., Luck, R. An Introduction to Sensor Signal Validation in Redundant Measurement Systems // IEEE Contr. Syst. Mag. – 1991. – Vol. 11, no. 2. – P. 44–49.
4. Chow, E.Y., Willsky, A.S. Issues in the Development of a General Algorithm for Reliable Failure Detection // Proceedings of the 19th Conf. on Decision & Control. – Albuquerque, 1980.
5. Основы технической диагностики. В 2-х книгах. Кн. 1. Модели объектов, методы и алгоритмы диагноза / Под ред. П.П. Пархоменко. – М.: «Энергия», 1976. – 464 с. [Osnovy tekhnicheskoy diagnostiki. V 2-h knigah. Kn. 1. Modeli ob'ek-tov, metody i algoritmy diagnoza / Pod red. P.P. Parhomenko. M.: «Energiya», 1976. – 464 s. (In Russian)]
6. Мозгалеvский А.В. Техническая диагностика (непрерывные объекты) (обзор) // Автоматика и телемеханика. – 1978. – № 1. – С. 145–166. [Mozgalevskii, A.V. Technological Diagnostics. Continuous Objects (Survey) // Automation and Remote Control. – 1978. – No. 1. – P. 145–166. (In Russian)]
7. Шумский А.Е., Жирабок А.Н. Методы и алгоритмы диагностирования и отказоустойчивого управления динамическими системами. – Владивосток: Изд. ДВГТУ, 2009. [Shumskij, A.E., Zhirabok, A.N. Metody i algoritmy diagnostirovaniya i otkazoustojchivogo upravleniya dinamicheskimi sistemami. – Vladivo-stok: Izd. DVGTU, 2009. (In Russian)]
8. Кулида Е.Л., Лебедев В.Г. Современные подходы к диагностике и прогнозированию технического состояния электромеханического привода летательного аппарата // Проблемы управления. – 2024. – № 3. – С. 3–19. [Kulida, E.L. and Lebedev, V.G. Modern Approaches to Prognostics and Health Management of an Aircraft Electromechanical Actuator // Control Sciences. – 2024. – No. 3. – P. 2–15.]
9. Diagnosis and Fault-tolerant Control 1: Data-Driven and Model-Based Fault Diagnosis Techniques / ed. by V. Puig and S. Simani. – Hoboken: John Wiley & Sons, 2021.
10. Мироновский Л.А., Соловьева Т.Н., Шинтяков Д.В. Оптимизация терминального контроля управляемых динамических систем // Информационно-управляющие системы. – 2019. – № 6 (103). – С. 12–21. [Mironovsky, L.A., Solov'eva, T.N., and Shintyakov, D.V. Fault Detection Pptimization for Controllable Dynamic Systems // Information and Control Systems. – 2019. – No. 6. – P. 12–21. (In Russian)]
11. Мироновский Л.А. Функциональное диагностирование динамических систем (обзор) // Автоматика и телемеханика. – 1980. – № 8. – С. 96–121. [Mironovsky, L.A. Functional Diagnosis of Dynamic Systems // Avtomatika I Telemekhanika. – 1980. – No. 8. – P. 96–121. (In Russian)]
12. Basseville, M. Detecting Changes in Signals and Systems: A Survey // Automatica. – 1988. – Vol. 24, no. 3. – P. 309–326.
13. Billings, S.A. Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains. – Hoboken: Wiley, 2013.
14. Issues of Fault Diagnosis for Dynamic Systems / ed. by R. J. Patton, P. M. Frank, R. N. Clark. – London: Springer-Verlag, 2000.
15. Fault Detection, Supervision and Safety for Technical Processes. Safeprocess'91 / ed. by R. Isermann, B. Freyermuth. – Oxford: Pergamon Press, 1992.
16. Чернодаров А.В. Контроль, диагностика и идентификация авиационных приборов и измерительных вычислительных комплексов. – М.: ООО «Научтехлитиздат», 2017. – 300 с. [Chernodarov, A.V. Kontrol', diagnostika i identifikaciya aviacionnyh priborov i izmeritel'nyh vychislitel'nyh kompleksov. – M.: ООО «Nauchtekhlitizdat», 2017. – 300 s. (In Russian)]
17. Jazwinski, A. H. Stochastic Processes and Filtering Theory. – New York: Academic Press, 1970.
18. Тотиев Д.А., Виниченко С.Н. Применение фильтра Калмана для обработки сигналов с измерительных устройств // Сб. науч. тр. кафедры автоматизации и промышленной электроники. – М.: Изд. РГУ им. А. Н. Косыгина. – 2024. – С. 165–168. [Totiev, D.A., Vinichenko, S.N. Primenenie fil'tra Kalmana dlya obrabotki signalov s izmeritel'nyh ustrojstv // Sb. nauch. tr. kafedry avtomatiki i promyshlennoj elektroniki. – M.: Izd. RGU im. A. N. Ko-sygina. – 2024. – P. 165–168. (In Russian)]
19. Буков В.Н., Бронников А.М., Попов А.С., Шурман В.А. Методы мониторинга технического состояния систем в целях управления их избыточностью. Ч. 1. Встроенные средства контроля и разбиение на классы // Проблемы управления. – 2025. – № 2. – С. 3–13. [Bukov, V.N., Bronnikov, A.M., Popov, A.S., Shurman, V.A. Technical Condition Monitoring Methods to Manage the Redundancy of Systems. Part I: Built-in Control and Partition into Classes // Control Sciences. – 2025. – No. 2. – P. 2–10.]
20. Chen, J., Patton, R. J. Robust Model-Based Fault Diagnosis for Dynamic Systems. – Norwell: Kluwer Academic Publishers, 1999.
21. Frank, P.M. Enhancement of Robustness on Observer-Based Fault Detection // Int. J. of Control. – 1994. – Vol. 59, no. 4. – P. 955–983.



22. Frank, P.M., Ding, X. Survey of Robust Residual Generation and Evaluation Methods in Observer-Based Fault Detection System // J. of Process Control. – 1997. – Vol. 7, no. 6. – P. 403–424.
23. Liu, G.P., Patton, R.J. Eigenstructure Assignment for Control System Design. – Chichester: John Wiley & Sons, 1998.
24. Zhirabok, A.N., Shumsky, A.Y., Suvorov, A.Y. Observer-Based Robust Fault Diagnosis: Logic-Dynamic Approach // Proceedings of the 10th International Conference on Informatics in Control, Automation and Robotics (ICINCO). – Reykjavik, Iceland, 2013. – P. 239–244.
25. Isermann, R. Estimation of Physical Parameters for Dynamic Processes with Application to an Industrial Robot // Int. J. of Control. – 1992. – Vol. 55. – P. 1287–1298.
26. Clark, R.N. Instrument Fault Detection // IEEE Trans. Aero. & Electronic Systems. – 1978. – Vol. 14, no. 3. – P. 455–465.
27. Willsky, A.S. A survey of design methods for failure detection in dynamic systems // Automatica. – 1976. – Vol. 12, no. 6. – P. 601–611.
28. Wünnenberg, J., Frank, P.M. Sensor Fault Detection via Robust Observer // System Fault Diagnosis, Reliability, and Related Knowledge-Based Approaches / ed. by S. Tzafestas, M. Singh, G. Schmidt. – Cham, Switzerland: Springer, 1987.
29. Frank, P.M. Advances in Observer-Based Fault Diagnosis // Proceedings of TOOLDIAG'93 Conference. – Toulouse, France, 1993. – P. 817–836.
30. Жирабок А.Н., Зуев А.В. Интервальные наблюдатели для оценивания неизвестных входов в дискретных стационарных динамических системах // Автоматика и телемеханика. – 2024. – № 8. – С. 3–19. [Zhirabok, A.N. and Zuev, A.V. Interval Observers for Estimating Unknown Inputs in Discrete Time-Invariant Dynamic Systems // Automation and Remote Control. – 2024. – Vol. 85, iss. 8. – P. 747–760.]
31. Beard, R.V. Failure Accommodation in Linear Systems through Self-Reorganisation: Technical Report MVT-71-1. – Cambridge: Massachusetts Institute of Technology, Man Vehicle Lab., 1971.
32. Jones, H.L. Failure Detection in Linear Systems : PhD Thesis. Cambridge: Massachusetts Institute of Technology, Department of Aeronautics, 1973.
33. Chen, J., Patton, R.J., Zhang, H.Y. Design of Unknown Input Observer and Robust Fault Detection Filters // Int. J. Control. – 1996. – Vol. 63, no. 1. – P. 85–105.
34. Gertler, J. Generating Directional Residuals with Dynamic Parity Equations // Proceedings of the IFAC/IMACS Symp. SAFEPROCESS'91. – Baden-Baden, 1991.
35. Patton, R.J., Chen, J. A Review of Parity Space Approaches to Fault Diagnosis for Aerospace Systems // AIAA J. of Guidance, Contr. & Dynamics. – 1994. – Vol. 17, no. 2. – P. 278–285.
36. Fault Diagnosis in Dynamic Systems, Theory and Application / ed. by R.J. Patton, P.M. Frank, R.N. Clark. – London: Prentice Hall, 1989.
37. Chow, E.Y., Willsky, A.S. Analytical Redundancy and the Design of Robust Detection Systems // IEEE Trans. Automatic Control. – 1984. – Vol. 29, no. 7. – P. 603–614.
38. Бритов Г.С., Мироновский Л.А. Метод избыточных переменных для контроля звеньев с дробно-рациональными передаточными функциями // Информ. материалы НС по проблеме «Кибернетика» АН СССР. – 1971. – № 3. [Britov, G.S., Mironovskij, L.A. Metod izbytochnykh peremennykh dlya kontrolya zven'ev s drobnoratsional'nymi peredatochnymi funktsiyami // Inform. materialy NS po probleme «Kibernetika» AN SSSR. – 1971. – No. 3. (In Russian)]
39. Игнатьев М.Б., Катермина Т.С. Метод избыточных переменных для контроля и коррекции вычислительных процессов в реальном времени // Труды СПИИРАН. – 2013. – Вып. 3 (26). – С. 234–252. [Ignat'ev, M.B., Katermina, T.S. Metod izbytochnykh peremennykh dlya kontrolya i korrektsii vychislitel'nykh processov v real'nom vremeni // Trudy SPIIRAN. – 2013. – Vyp. 3 (26). – S. 234–252. (In Russian)]
40. Жирабок А.Н., Кучер Д.Н. Метод повышения робастности при диагностировании // Вологодские чтения. – 2008. – № 69. – С. 6–9. [Zhirabok, A.N., Kucher, D.N. Metod povysheniya robastnosti pri diagnostirovani // Vologdinskie chteniya. – 2008. – No. 69. – P. 6–9. (In Russian)]
41. Emami-Naeini, A., Akhter, M., Rock, M. Effect of Model Uncertainty on Failure Detection: The Threshold Selector // IEEE Trans. on Automatic Control. – 1988. – Vol. 33, no. 2. – P. 1106–1115.
42. Frank, P.M., Ding, S.X., Köpper-Seliger, B. Current Developments in the Theory of FDI // SAFEPROCESS'00: Preprints of the IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, Vol. 1. – Budapest, Hungary. – London: Elsevier Ltd., 2000.
43. Zhou, K., Doyle, J. C., Glover, K. Robust and Optimal Control. – Upper Saddle River, NJ: Prentice Hall, 1996.

Статья представлена к публикации членом редколлегии
В.Г. Лебедевым.

Поступила в редакцию 10.11.2024,
после доработки 23.04.2025.
Принята к публикации 16.06.2025.

Буков Валентин Николаевич – д-р техн. наук, АО «Бортовые аэронавигационные системы», г. Москва,
✉ v_bukov@mail.ru,
ORCID iD: <https://orcid.org/0000-0002-5194-8251>

Бронников Андрей Михайлович – д-р техн. наук, МГТУ им. Н.Э. Баумана, ✉ bronnikov_a_m@mail.ru,
ORCID iD: <https://orcid.org/0009-0009-1216-3521>

Попов Александр Сергеевич – канд. техн. наук, ВУНЦ ВВС «Военно-воздушная академия им. проф. Н. Е. Жуковского и Ю. А. Гагарина», г. Воронеж, ✉ saga30@yandex.ru,

Шурман Владимир Александрович – АО «НИИ авиационного оборудования», г. Жуковский, ✉ shurman@niiao.ru

© 2025 г. Буков В. Н., Бронников А. М., Попов А. С., Шурман В. А.



Эта статья доступна по лицензии [Creative Commons «Attribution» \(«Атрибуция»\) 4.0 Всемирная](https://creativecommons.org/licenses/by/4.0/).

TECHNICAL CONDITION MONITORING METHODS TO MANAGE THE REDUNDANCY OF SYSTEMS.

PART II: Classical Models

V. N. Bukov*, A. M. Bronnikov**, A. S. Popov***, and V. A. Shurman****

*Airborne Aeronautical Systems, Moscow, Russia

**Bauman Moscow State Technical University, Moscow, Russia

***Zhukovsky–Gagarin Air Force Academy, Voronezh, Russia

****Institute of Aircraft Equipment, Zhukovsky, Russia

*✉ v_bukov@mail.ru, **✉ bronnikov_a_m@mail.ru, ***✉ saga30@yandex.ru, ****✉ shurman@niiio.ru

Abstract. Redundancy management of a technical system involves a monitoring procedure (control of the current state of its components) to reconfigure the system as needed. Consisting of four parts, this survey presents modern and newly developed technical condition monitoring methods for redundancy management. Part I was devoted to a general description of built-in control tools, voting schemes, and fidelity rules; control codes and program execution control methods were briefly covered. Part II of the survey considers diagnosis methods based on the classical modeling of the system diagnosed in the discrete time and frequency domains. The Chow–Willsky fault detection scheme, as well as the definition of a residual and its generation procedures in the diagnosis problem, are presented. The main model-based diagnosis methods using equation errors, observers, parity equations, and redundant variables are described. In conclusion, the robustness problem of diagnosis methods of the corresponding type is briefly discussed.

Keywords: technical condition monitoring, redundancy management, diagnosis, residual, analytical models of systems, equation error method, observers, Beard filter, fault detection via parity equations, redundant variable method, robustness of methods.

ОЦЕНКА ФУНДАМЕНТАЛЬНОЙ ЧАСТОТЫ ВАЛЬНО-ЛОПАСТНОГО РЯДА С ПРИМЕНЕНИЕМ ПРЕОБРАЗОВАНИЯ ГИЛЬБЕРТА И АВТОКОРРЕЛЯЦИИ[#]

О. В. Бабилов*, В. Г. Бабилов**

***Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

*✉ babikov.ov@phystech.edu, *✉ babikov@ipu.ru

Аннотация. Задача оценки фундаментальной частоты ряда гармоник встречается во многих областях науки и техники. Так, в задачах вибрационной диагностики требуется, например, оценить износ подшипников, что определяется по смещению основания ряда гармоник. При обработке аудиосигналов задача оценки основной частоты связана с автоматической настройкой инструментов. В задаче синтеза речи фундаментальная частота определяет высоту звука. При распознавании речи частота основного тона является важным информационным признаком. В радиотехнике эта задача решается в целях восстановления сигнала, фильтрации и декодирования. В биомедицинской инженерии при анализе ЭКГ, ЭЭГ, голоса, дыхания по основной частоте диагностируются патологии, например, аритмии. В задачах обнаружения и классификации морских судов важнейшим информационным критерием является основание вально-лопастного ряда. В данной работе предлагаются новые подходы для оценки фундаментальной частоты в условиях сильного шума. В целях снижения ошибок предлагается применять метод периодограмм, фильтрацию, автокорреляцию, преобразование Гильберта. Стоит также отметить, что качество оценки основной частоты ряда гармоник в условиях повышенного шума значительно возрастает при подборе оптимальных параметров: размера временного окна, параметров фильтрации, интервала спектра для автокорреляции, количества автокорреляций.

Ключевые слова: дискретное преобразование Фурье, автокорреляция, преобразование Гильберта, фундаментальная частота.

ВВЕДЕНИЕ

Конкретизация задачи в вопросе поиска фундаментальной частоты связана с различиями в типах сигналов и шумов, методах предобработки сигнально-шумовой смеси. Далее в работе рассматривается задача поиска основания вально-лопастного ряда (ОВЛР) создаваемого винтом морского судна, хотя предлагаемые методы применимы и в других инженерных дисциплинах. Одним из главных источников первичного гидроакустического поля морских судов являются гребные винты. Они формируют вибрации на двух ключе-

вых дискретных частотах: на частоте вращения вала (вальная частота) и на частоте, равной произведению вальной частоты на число лопастей винта (лопастная частота) [1, 2]. У современных судов вальные частоты обычно находятся в диапазоне 1–6 Гц, а лопастные – в пределах 6–24 Гц [3].

Из-за нелинейных эффектов, возникающих при излучении акустических волн, в низкочастотном спектре судового шума образуется совокупность гармоник с кратными частотами. Эти гармоники, амплитуда которых значительно превышает уровень окружающего шума, называются дискретными составляющими. Группу таких дискретных составляющих, расположенных на кратных частотах, называют звукорядом. Если источник этих гармоник – гребной винт, звукоряд называется вально-лопастным.

[#] Работа выполнена при частичной финансовой поддержке Российского научного фонда (грант № 23-19-00134).

В вально-лопастном звукоядре можно выделить два типа дискретных составляющих: вального и лопастного звукоядер. Первая дискретная составляющая вального звукоядра соответствует частоте вращения вала, которая напрямую связана со скоростью движения судна [2, 4]. Основная частота лопастного звукоядра определяется произведением вальной частоты на число лопастей винта. Таким образом, по анализу вально-лопастного звукоядра можно получить ценную информацию о конструкции судна, включая количество лопастей гребного винта, что активно используется в системах распознавания морских целей, основанных на анализе спектра сигнала [5].

Современные гидроакустические системы позволяют проводить анализ таких сигналов с высокой точностью. Однако, несмотря на развитые алгоритмы обработки данных, окончательное решение в большинстве случаев принимает оператор [6]. Определение количества лопастей винта осуществляется на основе представленных оператору параметров спектра вально-лопастного звукоядра.

Среди методов обработки сигналов выделяют вейвлет-анализ, который позволяет обнаруживать гидроакустические сигналы в виде звукоядра и измерять основную частоту вального звукоядра [4]. Другим подходом является спектральный анализ с последовательным выделением отдельных дискретных составляющих и формированием звукоядер [3].

Для удобства оператора результаты узкополосного частотного анализа обычно представляют в двух формах: график спектра сигнала, отображающий расположение гармоник; таблица параметров обнаруженных дискретных составляющих и их характеристик. При анализе графика спектра оператор визуально идентифицирует гармоники, чтобы выделить среди них основные вальную и лопастную частоты. Вальная частота, как правило, является первой по порядку дискретной составляющей, а лопастная – одной из последующих с максимальной амплитудой. Кроме того, таблица параметров дискретных составляющих предоставляет оператору числовые значения частот, что повышает точность идентификации и снижает вероятность ошибки.

Современные исследования направлены на разработку автоматизированных алгоритмов, способных снизить зависимость анализа от человеческого фактора. Наиболее традиционным инструментом для анализа акустического шума является преобразование Фурье (англ. *Fast Fourier Transform*, FFT), которое позволяет выделять основные гармониче-

ские составляющие сигнала. Однако FFT имеет ограниченную разрешающую способность, особенно при низком уровне сигнала и высоком уровне шума. Для преодоления этих ограничений применяются методы периодограммы и автокорреляционного анализа. Например, метод Уэлча уменьшает разброс оценок спектральной плотности мощности [7], а методы сверхвысокого разрешения MDVR (англ. *Minimum Variance Distortionless Response*), MUSIC (англ. *MUltiple Signal Classification*), ESPRIT (англ. *Estimation of Signal Parameters via Rotational Invariant Techniques*) обеспечивают повышенную разрешающую способность [8].

Автокорреляция эффективна при анализе слабых сигналов и позволяет выделять закономерности даже при сильном шуме. Также активно используется анализ огибающей спектра, который применяется для выделения характеристик лопастного шума. В работе [9] предложена методика адаптивного спектрального анализа огибающей, позволяющая учитывать влияние флуктуаций в спектре.

В настоящей работе предлагаются несколько подходов при исследовании смеси гармонического звукоядра и шума, которые в комплексе позволяют производить оценку основания гармонического ряда даже при малом отношении сигнал/шум по одному временному окну. В качестве основы предлагается использовать логарифм энергетического спектра исследуемого сигнала. Далее производится оценка огибающей логарифма энергетического спектра итерационным усреднением по трем точкам. Как показано в статье, такой метод весьма близок к свертке спектра с функцией Гаусса, но при этом стоит сказать, что усреднение по трем точкам позволяет в конечном итоге производить оценку основания гармонического ряда с меньшими ошибками. Далее анализируется псевдоспектр, который представляет собой разницу логарифма энергетического спектра и его огибающей.

Метод последовательных автокорреляций не всегда дает возможность верно оценить основание гармонического ряда. На результат влияют и плотность спектральных линий, и значение основания звукоядра, а также уровень шума. С целью повышения чувствительности метода предлагается до применения автокорреляции слегка «размыть» дискрету гармоник, с этой целью псевдоспектр чуть сглаживается.

После нескольких последовательных автокорреляций при помощи преобразования Гильберта вычисляется так называемая кепстральная фаза или сафе (см. пояснение в § 1). Дело в том, что

после последовательных автокорреляций псевдоспектра результат напоминает убывающую гармонику во временной области, но при этом полученный ряд значений относится к частотной области. Поэтому, применяя преобразование Гильберта к автокорреляции псевдоспектра, на деле получают оценку кепстральной фазы, которая при делении на частоту, к которой относится эта фаза, дает оценку основания звукоряда.

С развитием методов искусственного интеллекта стали популярны нейросетевые модели, способные классифицировать суда по их акустическим сигналам. В исследованиях [10, 11] рассмотрены сверточные нейронные сети (англ. *Convolutional Neural Network, CNN*), обученные на спектрограммах шумов. Подходы на основе машинного обучения позволяют повысить точность определения значений параметров винта, однако требуют больших объемов данных для обучения. Несмотря на прогресс в области автоматизации, участие оператора остается важным элементом анализа, особенно в сложных акустических условиях, где возможны помехи и ложные срабатывания.

Метод DEMON (англ. *Demodulation of Envelope Modulation On Noise*) специально разработан для обнаружения модуляций, возникающих от огибающей кавитации гребного винта. При помощи DEMON можно выделить кавитационный шум из общего спектра сигнала и определить количество валов, частоту вращения вала, количество лопастей винта. В биоакустике при

помощи данного метода осуществляется анализ сигналов китов, дельфинов, других животных. Ограничение данного метода заключается в необходимости выбора шумовой полосы, что требует хороших навыков работы оператора-гидроакустика. Кроме того, низкая помехоустойчивость является еще одним недостатком этого метода. Предлагаемый в настоящей работе алгоритм, в составе которого используются сглаживание логарифма спектра мощности по трем точкам, предварительная фильтрация псевдоспектра, несколько последовательных автокорреляций и преобразование Гильберта, позволяет значительно увеличить помехоустойчивость при оценке ОВЛР.

1. ПОСТАНОВКА ЗАДАЧИ И ПОДРОБНОЕ ОПИСАНИЕ ЭТАПОВ ОЦЕНКИ ФУНДАМЕНТАЛЬНОЙ ЧАСТОТЫ

Сформулируем основную задачу настоящей работы. Пусть $\{x_k\}_{k=1}^N$ – дискретный ряд аудиозаписи шума морского судна во временном окне размера T , количество отсчетов в одном окне N , $k \in [1, N]$, $N = T f_s$, где T – длина окна в секундах; f_s – частота дискретизации. Необходимо построить алгоритм оценки ОВЛР f_0 по одному временному окну, а также определить оптимальные параметры этого алгоритма.

Далее представлено последовательное описание структурных элементов алгоритма оценки ОВЛР (рис. 1).

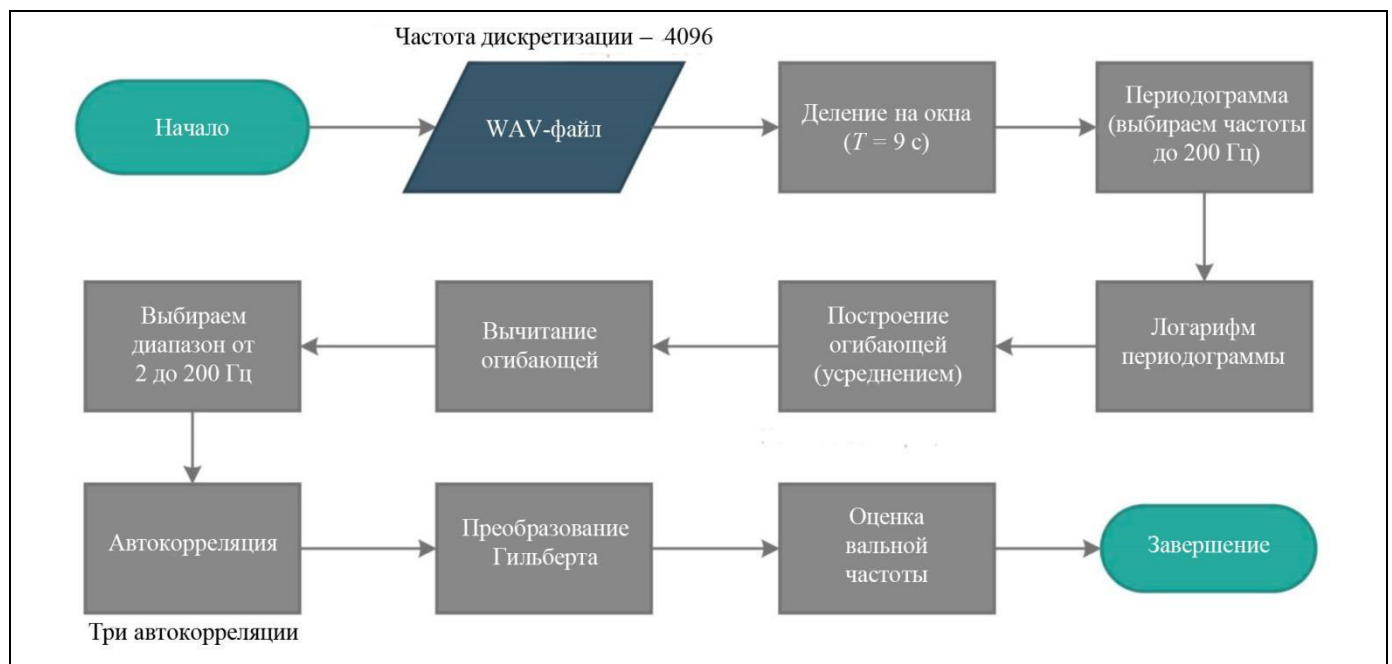


Рис. 1. Структура алгоритма определения ОВЛР

Построение периодограммы

Периодограмма вычисляется с помощью дискретного преобразования Фурье (англ. *Discrete Fourier Transform*, DFT) для сигнала x_k длины N :

$$X_j = \sum_{k=0}^{N-1} x_k e^{-i \frac{2\pi}{N} jk},$$

где X_j – комплексное значение спектра на j -й частоте ($j=0, 1, \dots, N/2$), соответственно, значение частоты в Гц вычисляется так: $f_j = j \frac{f_s}{N}$.

Далее в работе алгоритма используется только диапазон частот $f_{\min} \leq f_j \leq f_{\max}$. Ограничение по частотам связано с тем, что на частотах выше $f_{\max}$ отсутствуют составляющие гармонического ряда, а на частотах ниже $f_{\min}$ присутствует сильный шум, следовательно, $N \frac{f_{\min}}{f_s} < j < N \frac{f_{\max}}{f_s}$.

Логарифмирование периодограммы

При логарифмировании энергетического спектра сигнала исчезают резкие скачки, которые связаны с гармониками вально-лопастного ряда, что, в свою очередь, позволяет выделить огибающую при сглаживании:

$$S_j^{(0)} = \ln |X_j|^2.$$

Итеративное сглаживание спектра

Построение огибающей спектра достигается методом усреднения по трем точкам в M итерациях:

$$S_j^{(m)} = \frac{S_{j-1}^{(m-1)} + S_j^{(m-1)} + S_{j+1}^{(m-1)}}{3}, \quad (1)$$

где j – индекс элемента спектрального ряда

$$\left(N \frac{f_{\min}}{f_s} < j < N \frac{f_{\max}}{f_s} \right); \quad m \text{ – номер итерации}$$

($m \in [1, M]$); M – количество итераций.

Далее требуется уточнить, что границы спектра вносят некоторые коррективы в выражение (1):

$$S_j^{(m)} = \begin{cases} S_j^{(m-1)} & \text{(для граничных точек),} \\ \frac{S_{j-1}^{(m-1)} + S_j^{(m-1)} + S_{j+1}^{(m-1)}}{3} & \text{(для внутренних точек).} \end{cases} \quad (2)$$

Определение 1. Итеративный алгоритм сглаживания ряда, заданный выражением (2), будем называть сглаживанием по трем точкам порядка M . ♦

Коррективы в итоговом алгоритме сглаживания по трем точкам (2), учитывающие влияние граничных точек, не оказывают значительного влияния на результат сглаживания спектра при $M \ll N$, поэтому далее в целях упрощения доказательств в математических выкладках используется выражение (1).

Заметим, что M итераций в выражении (1) практически равносильны взвешиванию элементов исходного ряда по функции Гаусса со стандартным отклонением $\sigma \approx \sqrt{\frac{2M}{3}} \approx$

$\approx 0,820467 \sqrt{M}$. Примеры для $M=5$ и $M=50$ приведены на рис. 2 (см. пример для $M=1, \dots, 4$ в таблице).

Весовые коэффициенты алгоритма сглаживания по трем точкам в зависимости от номера итерации

m	$j-4$	$j-3$	$j-2$	$j-1$	j	$j+1$	$j+2$	$j+3$	$j+4$
1				$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$			
2			$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$		
3		$\frac{1}{27}$	$\frac{3}{27}$	$\frac{6}{27}$	$\frac{7}{27}$	$\frac{6}{27}$	$\frac{3}{27}$	$\frac{1}{27}$	
4	$\frac{1}{81}$	$\frac{4}{81}$	$\frac{10}{81}$	$\frac{16}{81}$	$\frac{19}{81}$	$\frac{16}{81}$	$\frac{10}{81}$	$\frac{4}{81}$	$\frac{1}{81}$
					...				

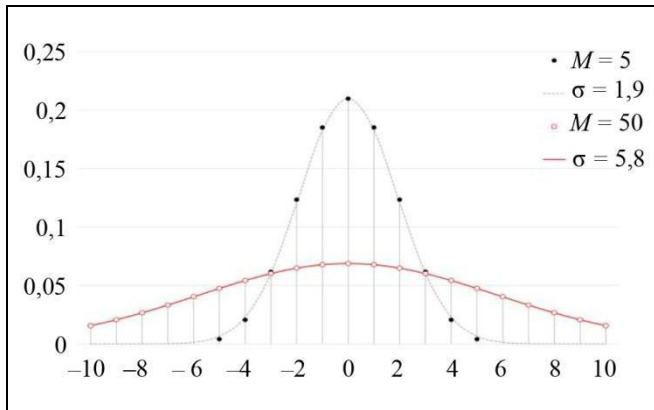


Рис. 2. Сравнение соответствующих весов элементов массива, полученных согласно формуле (1) при $M = 5$ и $M = 50$ и функции Гаусса ($\sigma = \sqrt{2M/3}$). По оси абсцисс указаны отклонения индекса элемента исходного ряда от индекса j при оценке весового коэффициента

Сформулируем следующую лемму.

Лемма 1 (вычисление весовых коэффициентов). M последовательных итераций (1) порождают весовые коэффициенты, приближенные оценки которых могут быть получены с помощью функции Гаусса с параметром

$$\sigma = \sqrt{\frac{2M}{3}} \quad (M \gg 1):$$

$$S_j^{(M)} = \sum_{k=-M}^M \left[\frac{S_{j+k}^{(0)}}{\sqrt{4\pi M/3}} \exp\left(-\frac{k^2}{4M/3}\right) \right]. \quad (3)$$

Доказательство. Доказательство того, что весовые коэффициенты, полученные посредством последовательных итераций, могут быть оценены функцией Гаусса, носит технический характер. Покажем, что при M итерациях веса при усреднении по трем точкам (1) аппроксимируются функцией Гаусса с параметром $\sigma = \sqrt{\frac{2M}{3}}$.

Пусть M достаточно большое. Рассмотрим итерацию $M+1$, тогда при условии аппроксимации весов в алгоритме усреднения по трем точкам на шаге M функцией Гаусса с параметром $\sigma_M^2 = \alpha M$ имеем следующее при $j \ll M$:

$$\begin{aligned} & \frac{1}{3\sigma\sqrt{2\pi}} \left(e^{-\frac{(j-1)^2}{2\sigma^2}} + e^{-\frac{(j)^2}{2\sigma^2}} + e^{-\frac{(j+1)^2}{2\sigma^2}} \right) \approx \\ & \approx \frac{1}{3\sigma\sqrt{2\pi}} \left(1 - \frac{(j-1)^2}{2\sigma^2} + 1 - \frac{(j)^2}{2\sigma^2} + 1 - \frac{(j+1)^2}{2\sigma^2} \right) = \\ & = \frac{1}{\sigma\sqrt{2\pi}} \left(1 - \frac{1}{2\sigma^2} \left(j^2 + \frac{2}{3} \right) \right) \approx \frac{1}{\sigma\sqrt{2\pi}} \left(1 + \frac{1}{3\sigma^2} \right) e^{-\frac{j^2}{2\sigma^2}}. \end{aligned}$$

Коэффициент при экспоненте на шаге $M+1$ должен соответственно определяться новым параметром $\sigma_{M+1}^2 = \alpha(M+1)$, поэтому

$$\sigma_{M+1}^2 = \alpha(M+1) = \alpha M + \alpha \approx \alpha M \left(1 + \frac{1}{3\alpha M} \right)^2 \approx \alpha M + \frac{2}{3}.$$

Откуда следует, что $\alpha = \frac{2}{3}$ и, соответственно,

$$\sigma = \sqrt{\frac{2M}{3}}. \quad \blacklozenge$$

Сформулируем следующие леммы.

Лемма 2 (сглаживание спектра при помощи свертки). Приближенная оценка результата сглаживания при помощи M последовательных итераций (1), примененных к логарифму энергетического спектра сигнала во временном окне T , может быть получена посредством свертки логарифма энергетического спектра сигнала и функции Гаусса с параметром

$$\sigma = \frac{1}{T} \sqrt{\frac{2M}{3}} \quad (M \gg 1):$$

$$S_j^{(M)} = \sum_{k=-M}^M \left[S_{j+k}^{(0)} \sqrt{\frac{3}{\pi}} M^{\frac{1}{2}} \exp\left(-\frac{3k^2 T^2}{4M}\right) \right]. \quad (4)$$

Альтернативная запись выглядит следующим образом:

$$S^{(M)}(f) = S^{(0)}(f) * N(f, \sigma). \quad (5)$$

Доказательство. При увеличении размера временного окна T плотность спектральных линий возрастает, а стандартное отклонение – параметр функции Гаусса, по которой согласно лемме 1 (см. формулу (3)) осуществляется оценка весовых коэффициентов, уменьшается пропорционально размеру окна T . Соблюдая условие нормировки функции Гаусса, получаем искомое выражение (4). $\blacklozenge$

Лемма 3 (сглаживание спектра при помощи кепстра и функции Гаусса). Приближенная оценка результата сглаживания при помощи M последовательных итераций (1), примененных к логарифму энергетического спектра сигнала во временном окне T , может быть получена посредством преобразования Фурье произведения кепстра этого сигнала $K(\tau) = F^{-1}[S^{(0)}(f)]$ на функцию Гаусса с параметром $\sigma^* = T \sqrt{\frac{3}{2M}}$ ($M \gg 1$):

$$S^{(M)} = F[K(\tau) \cdot N(\tau, \sigma^*)].$$

Доказательство. Действительно, так как выражение (5) представляет из себя свертку логарифма

энергетического спектра сигнала и функции Гаусса, то согласно известной формуле о преобразовании Фурье произведения двух функций ($F[f \cdot g] = F[f] * F[g]$) имеем, что представленное выражение есть не что иное как преобразование Фурье от произведения обратного преобразования Фурье логарифма энергетического спектра (что есть кепстр) и обратного преобразования Фурье от функции Гаусса в частотной области с параметром $\sigma = \frac{1}{T} \sqrt{\frac{2M}{3}}$ (что также является функцией Гаусса, но во временной области с параметром $\sigma^* = T \sqrt{\frac{3}{2M}}$), см. пример на рис. 3, а. ♦

Лемма 4 (выделение линейчатого спектра). Приближенная оценка линейчатого спектра сигнала $\hat{S}(f) = S^{(0)}(f) - S^{(M)}(f)$ (ранее эта разность была определена как псевдоспектр) может быть получена следующими способами:

– посредством вычисления свертки логарифма энергетического спектра сигнала с разностью дельта-функции $\delta(f)$ и функции Гаусса:

$$\hat{S}(f) = S^{(0)}(f) * (\delta(f) - N(f, \sigma)),$$

где $\delta(f) \approx N(f, \sigma_\delta)$, $\sigma_\delta \ll \sigma$, причем σ_δ определяет точность линейчатого спектра, а σ определяет степень сглаженности огибающей спектра.

– посредством преобразования Фурье произведения кепстра исходного сигнала и разности константы и функции Гаусса:

$$\hat{S}(f) = F[K(\tau) \cdot (\Psi - N(\tau, \sigma^*))],$$

где кепстр $K(\tau) = F^{-1}[S^{(0)}(f)]$, $\Psi = F^{-1}[\delta(f)]$ ($\delta(f) \approx N(f, \sigma^\delta)$).

Доказательство. Результат, сформулированный в лемме 4, является следствием леммы 2 и леммы 3, а также следствием аддитивности преобразования Фурье и формулы преобразования Фурье для дельта-функции:

$$\begin{aligned} \hat{S}(f) &= S^{(0)}(f) - S^{(M)}(f) = \\ &= S^{(0)}(f) - S^{(0)}(f) * N(f, \sigma) = \\ &= S^{(0)}(f) * (\delta(f) - N(f, \sigma)). \end{aligned}$$

В свою очередь, применяя формулу для преобразования Фурье произведения двух функций, получаем:

$$\begin{aligned} \hat{S}(f) &= S^{(0)}(f) * (\delta(f) - N(f, \sigma)) = \\ &= F[F^{-1}[S^{(0)}(f)] \cdot F^{-1}[(\delta(f) - N(f, \sigma))]] = \\ &= F[K(\tau) \cdot (\Psi - N(\tau, \sigma^*))]. \end{aligned}$$

При реализации численного расчета в случае дискретного преобразования Фурье оценка дельта-функции может быть получена из функции Гаусса с малым параметром $\sigma_\delta \ll \sigma$, с соответствующим условием нормировки (сумма всех весов, полученных при помощи аппроксимации функцией Гаусса, должна быть равна единице), значение $\Psi = F^{-1}[\delta(\tau)]$, см. пример на рис. 3, б. ♦

Стоит напомнить, что при применении методов, относящихся к определению кепстров, учитываются функции, которые можно рассматривать как спектры логарифмических спектров. По существу, понятие кепстра было введено уже в 1963 г.

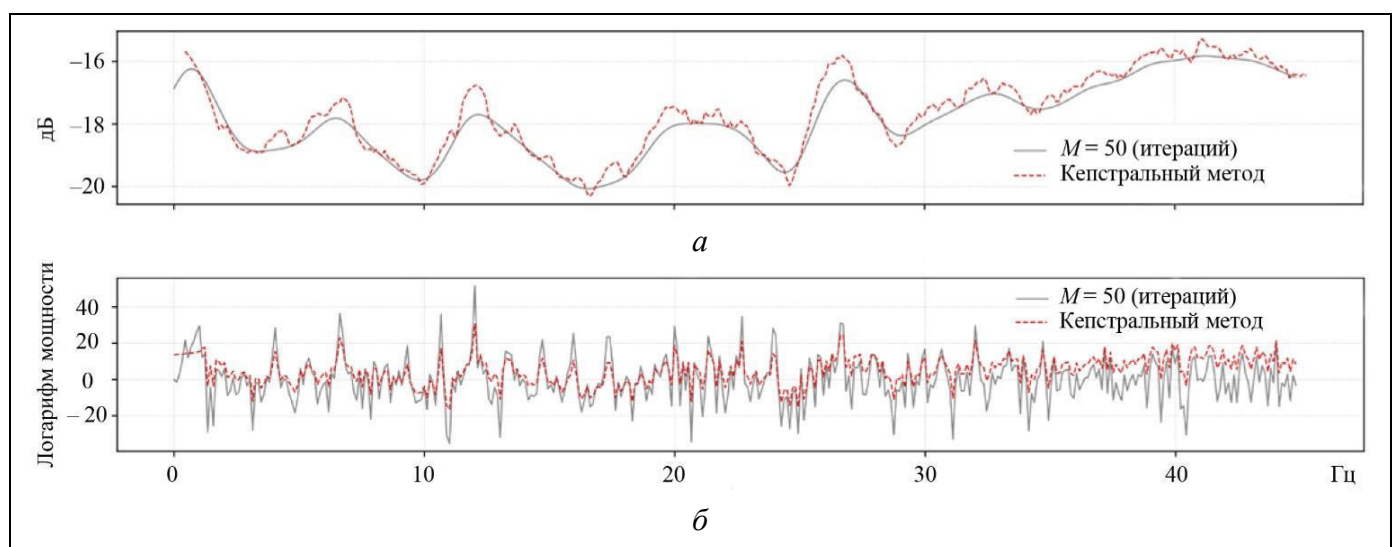


Рис. 3: а – оценка огибающей спектра (лемма 3), б – центрированный логарифмический спектр мощности (лемма 4). Оценки получены двумя методами: итеративное сглаживание спектра и вычитание огибающей (сплошные линии), оценка огибающей и оценка центрированного спектра посредством кепстрального метода (пунктирные линии)



Кепстр мощности определяют как «спектр мощности логарифмического спектра мощности» [12]. Кепстр мощности был предложен в качестве более эффективной альтернативы автокорреляционной функции при обнаружении эха в сигналах.

Поскольку соответствующая функция по определению отображала спектр спектра, автор работы [13] воспользовался терминологической аналогией и, согласно термину «спектр», придал этой функции название кепстр¹.

Однако самая важная особенность кепстра заключается не в том, что он представляет спектр спектра, а в логарифмическом преобразовании исходного и подвергаемого дальнейшей обработке спектра. Отметим, что автокорреляционную функцию, определяемую на основе собственного спектра мощности путем обратного преобразования Фурье, также можно рассматривать как «спектр спектра». По существу, используемое в настоящее время определение кепстра определяет кепстр мощности как «обратное преобразование Фурье логарифмического спектра мощности». Различие между этим определением и определением автокорреляционной функции заключается лишь в логарифмическом преобразовании исходного спектра.

Применение кепстров мощности в области исследования вально-лопастного ряда основано на обеспечиваемой кепстрами возможности обнаружения периодичностей спектров, например, серий равномерно распределенных гармоник. С точки зрения применения в упомянутой области важное преимущество кепстров связано с их малой зависимостью от путей распространения исследуемых сигналов, в том числе путей от источников к точкам замера.

Вычитание тренда и сужение частотного диапазона

Далее в целях выделения линейчатого спектра, который содержит важную информацию о вально-лопастном ряде, предлагается вычитать из логарифма спектра мощности сигнала его огибающую. Тогда искомым центрированный спектральный ряд получается следующим образом:

$$\hat{S}_j = \hat{S}(f_j) = S^{(0)}(f_j) - S^{(M)}(f_j) = S_j^{(0)} - S_j^{(M)}.$$

Затем определяется интервал частот $f_j \in [f_{\min}, f_{\max}]$. Соответственно, для индекса j этих частот справедливо следующее: $Tf_{\min} < j < Tf_{\max}$.

Последовательные автокорреляции

Применение нескольких последовательных автокорреляций к спектру (точнее к центрированному логарифму спектральной мощности) – это метод, который здесь применяется для дальнейшего сглаживания, улучшения выделения закономерностей и для более точного извлечения полезной информации из спектра.

Каждое вычисление автокорреляции усредняет информацию, что помогает удалить шумовые компоненты и краткосрочные флуктуации. Выполнение нескольких последовательных автокорреляций усиливает этот эффект. Шумовые компоненты часто имеют короткий период или случайный характер, а автокорреляция помогает выделить регулярные, повторяющиеся элементы. Если спектр сигнала имеет сложную структуру с несколькими периодическими компонентами (в спектре присутствует несколько гармонических рядов), выполнение нескольких автокорреляций позволяет более точно выделить эти компоненты. С каждым циклом автокорреляции спектра выделяются более устойчивые гармонические ряды, а неустойчивые гармонические ряды и разного рода шумы сглаживаются, поэтому несколько последовательных автокорреляций позволяют выделить наиболее устойчивую компоненту шума гребного винта – ОВЛР.

При анализе вально-лопастного ряда несколько последовательных шагов автокорреляции логарифма спектра мощности позволяют лучше выделить низкочастотную составляющую (частоту вала) на фоне более высокочастотной составляющей (лопастной частоты). При этом нужно иметь в виду, что при применении автокорреляции к спектру низкие и высокие частоты как бы меняются местами: низкочастотное вращение вала проявляется частыми пиками, а лопастная частота – более редкими пиками, и после каждой следующей автокорреляции вальная частота проявляется все сильнее.

Повторная автокорреляция спектра вычисляется уже для сглаженного результата из первой автокорреляции спектра, что дает дополнительное сглаживание. Поскольку шумовые компоненты в первой автокорреляции ослабляются, повторное усреднение их во второй автокорреляции приводит к еще большему ослаблению как случайных

¹ Аналогичным образом возникли и термины «квифренция» (*quefreny*), «рагмоника» (*rahmonic*), «лифтр» (*lifter*), «гамнитуда» (*gamnitude*) и «сафе» (*saphe*), основанные на аналогии с английскими терминами для частоты (*frequency*), гармоники (*harmonic*), фильтра (*filter*), модуля (*magnitude*) и фазы (*phase*). Относительно часто используются и термины, относящиеся к лифтру (лифтрация, пролифтрованный и т. п.) и указывающие на процесс фильтрации в кепстральной области.

выбросов, так и тех компонент, которые не образуют гармонических рядов. При этом существует ограничение: большое количество автокорреляций в конце концов сгладит даже полезную информацию об устойчивых, регулярных закономерностях, которые имеют долгосрочную структуру, связанную с ОВЛР.

Определим автокорреляционную функцию центрированного логарифмического спектра мощности $\hat{S}_j$ порядка $p = 0$ следующим образом:

$$C^{(0)}(f_k) = C_k^{(0)} = \sum_{j=Tf_{\min}}^{Tf_{\max}-k} \hat{S}_j \hat{S}_{j+k}.$$

Далее, определим $C_k^{(p)}$ как автокорреляционную функцию центрированного логарифмического спектра мощности $\hat{S}_j$ порядка p :

$$C^{(p)}(f_k) = C_k^{(p)} = \sum_{j=Tf_{\min}}^{Tf_{\max}-k} C_j^{(p-1)} C_{j+k}^{(p-1)},$$

где $p \in [1, P]$, $P + 1$ – количество последовательных автокорреляций.

Преобразование Гильберта

Современные методы теории аналитического сигнала [14] позволяют выделить (демодулировать) из колебательного процесса его мгновенную амплитуду (огибающую), мгновенную фазу и мгновенную частоту. Для получения этих мгновенных функций необходимо с помощью *интегрального преобразования Гильберта* преобразовать исходный процесс $x(t)$, заданный на некотором интервале, в сопряженный процесс $\hat{x}(t)$ [5]:

$$\hat{x}(t) = \mathcal{H}\{x(t)\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau.$$

Аналитический сигнал можно записать как

$$x_a(t) = x(t) + i\hat{x}(t).$$

Легко показать, что функция $\sin \omega_0 t$ является преобразованием Гильберта функции $\cos \omega_0 t$. Поэтому аналитический сигнал, соответствующий $\cos \omega_0 t$, имеет вид:

$$x_a(t) = \cos \omega_0 t + i \sin \omega_0 t = \exp(i\omega_0 t).$$

Аналитический сигнал общего вида удобно представлять в экспоненциальной форме как

$$x_a(t) = |x_a(t)| \exp[i\Phi(t)],$$

где

$$\begin{aligned} |x_a(t)| &= [x^2(t) + \hat{x}^2(t)]^{1/2}, \\ \Phi(t) &= \arctg[\hat{x}(t) / x(t)]. \end{aligned} \quad (6)$$

Теперь положим $\Phi(t) = \omega_0 t + \varphi(t)$ и запишем:

$$x_a(t) = |x_a(t)| \exp[i\varphi(t)] \exp(i\omega_0 t) = \theta(t) \exp(i\omega_0 t).$$

Комплексная огибающая $\theta(t)$ получается удалением комплексного множителя, связанного с несущей, из аналитического сигнала:

$$\theta(t) = x_a(t) \exp(-i\omega_0 t) = |x_a(t)| \exp[i\varphi(t)].$$

Если $\theta(t)$ – узкополосная относительно $\omega_0 / 2\pi$ функция, то она будет обладать свойствами, которые интуитивно связываются с понятием огибающей.

Физический смысл интегрального преобразования Гильберта для сигнала, представленного в частотной области, заключается в фазовом сдвиге всех спектральных составляющих исходного сигнала на $\pi / 2$. Двойное преобразование Гильберта приводит к исходному процессу, но только с обратным знаком, т. е. осуществляет сдвиг исходного сигнала на π .

Применим преобразование Гильберта к $C^{(p)}(f_k)$ для получения *аналитической автокорреляции*:

$$\tilde{C}^{(p)}(f_k) = C^{(p)}(f_k) + i\mathcal{H}\{C^{(p)}(f_k)\}.$$

Далее здесь кепстральная фаза определяется как аргумент аналитической автокорреляции:

$$\begin{aligned} \varphi_k &= \arg \tilde{C}^{(p)}(f_k) = \arg \left(C^{(p)}(f_k) + i\mathcal{H}\{C^{(p)}(f_k)\} \right), \\ \varphi_k &= \arctg \left(\frac{\mathcal{H}\{C^{(p)}(f_k)\}}{C^{(p)}(f_k)} \right). \end{aligned} \quad (7)$$

Остановимся на этом подробнее и сформулируем следующее

Определение 2. Будем называть *кепстральной фазой* значение, вычисляемое посредством выражения (7). ♦

Далее частоту, для которой вычисляется кепстральная фаза, будем обозначать f_H , подчеркивая тем самым, что она (сафе) вычисляется посредством преобразования Гильберта.

Лемма 5 (основание гармонического ряда и преобразование Гильберта). Пусть задан некоторый сигнал $x(t)$ с непостоянной амплитудой и медленно меняющейся в течение всего периода наблюдения T частотой. Тогда средний период колебаний сигнала $\bar{T}_0$ за время t может быть оценен посредством следующего выражения:

$$\bar{T}_0 = 2\pi t \left[\arctg \left(\frac{\mathcal{H}\{x(t)\}}{x(t)} \right) \right]^{-1}, \quad 0 < t < T.$$



Доказательство носит технический характер и легко может быть получено из выражения (6).

Следствием этой леммы является следующий результат. Представим вместо сигнала $x(t)$ (о котором идет речь в лемме 5) коррелограмму $C^{(P)}$ логарифмического энергетического спектра сигнально-шумовой смеси, содержащей сигнал (гармонический ряд с основанием f_0) и некоторый шум, причем отношение энергии сигнала (гармонического ряда) к энергии шума достаточно высокое. Тогда для оценки основания гармонического ряда справедливо следующее выражение:

$$f_0 = 2\pi f_H \left[\arctg \left(\frac{\mathcal{H}\{C^{(P)}(f_H)\}}{C^{(P)}(f_H)} \right) \right]^{-1}, \quad (8)$$

где $f_{\min} < f_H < f_{\max}$.

Причем стоит отметить, что частоту f_H следует выбирать в зависимости от конкретной задачи и условий. Так, если требуется, чтобы метод работал в условиях сильного шума, то рекомендуется выбирать значение этой частоты в диапазоне $0,1f_{\max} < f_H < 0,2f_{\max}$ (коррелограмма «ломается» на правом конце вследствие высокой шумовой компоненты). Если же требуется оценивать основание вально-лопастного ряда с высокой точностью в условиях слабой зашумленности, то следует выбирать $f_H \approx f_{\max}$. Нетрудно показать, что оценка ОВЛР посредством выражения (8) представляет усреднение частотных разностей пиков автокорреляции, а это, в свою очередь, определяет точность оценки $T^{-1} \frac{f_0}{f_H}$, что на

практике составляет тысячные доли герца (например, когда размер временного окна $T = 10$ с, основание вально-лопастного ряда $f_0 \approx 1$ Гц, а диапазон поиска мод гармонического ряда ограничен частотой $f_{\max} = 100$ Гц).

В целом отметим, что метод определения основания гармонического ряда посредством выражения (8) дает меньше ошибок с увеличением значения сигнал/шум (англ. *Signal-to-Noise Ratio*, SNR) и релевантные оценки можно получить только выше некоторого порогового значения $\text{SNR} > \text{SNR}_0$. Далее покажем, как дополнительно можно снизить этот порог.

На рис. 4 показан результат численного эксперимента (сплошная линия (без сглаживания): $\Omega = (1 + \exp(-0,4364 \cdot \text{SNR} - 0,8545))^{-1}$, пунктирная линия: $\Omega = (1 + \exp(-0,4212 \cdot \text{SNR} - 2,3633))^{-1}$). Каждая точка получена по 500 кейсам. Каждый кейс был получен следующим образом: генерация временного ряда размера $N = 4096$ посредством

суммирования сигнала (гармонического ряда) и белого шума; размер временного окна $T = 1$ с; количество дискрет гармонического ряда $d = 15$; энергии дискрет гармонического ряда равны друг другу; частотная разница между дискретами $f_0 = 60$ Гц; отношение сигнал/шум в ходе эксперимента изменялось от -12 дБ до 7 дБ.

Причем выяснилось, что предварительное сглаживание полученного сгенерированного псевдоспектра позволяет уменьшить пороговые значения отношения сигнал/шум с нуля до -5 дБ. На рис. 4, а показано, что оптимум достигается посредством 40 итераций сглаживания псевдоспектра по трем точкам, что для этого эксперимента эквивалентно свертке псевдоспектра с функцией Гаусса ($\sigma \approx 5$ Гц). На рис. 4, б представлен сравнительный анализ безошибочности оценок основания гармонического ряда при отсутствии сглаживания (нижняя кривая) и при сглаживании посредством 40 итераций (верхняя кривая). В реальных записях, как правило, псевдоспектр получается более или менее «размытым» и дополнительное сглаживание не всегда требуется.

На рис. 5 представлены гистограммы оценок основания гармонического ряда для различных отношений энергии SNR гармонического ряда и белого шума. Количество сгенерированных сигнально-шумовых смесей для построения каждой гистограммы равно одной тысяче. Частотная разница между гармониками равна 60 Гц. Предварительного сглаживания псевдоспектра не производилось.

На рис. 6 представлены гистограммы оценок основания гармонического ряда для уровня $\text{SNR} = -3$ дБ. Количество сгенерированных сигнально-шумовых смесей для построения каждой гистограммы равно 5000. Частотная разница между гармониками равна 60 Гц. На рис. 6, а рассмотрен вариант без сглаживания, на рис. 6, б рассмотрен вариант с предварительным сглаживанием (100 итераций по трем точкам) псевдоспектра сигнально-шумовой смеси.

2. КРАТКАЯ ЗАПИСЬ АЛГОРИТМА ОЦЕНКИ ОВЛР

В результате проверки алгоритма оценки ОВЛР на реальных звукозаписях шумов морских судов была подтверждена эффективность итерационного метода сглаживания по трем точкам при построении огибающей спектра. Также была подтверждена эффективность метода оценки ОВЛР на основе свертки логарифма энергетического спектра сигнала с разностью двух гауссовских функций.

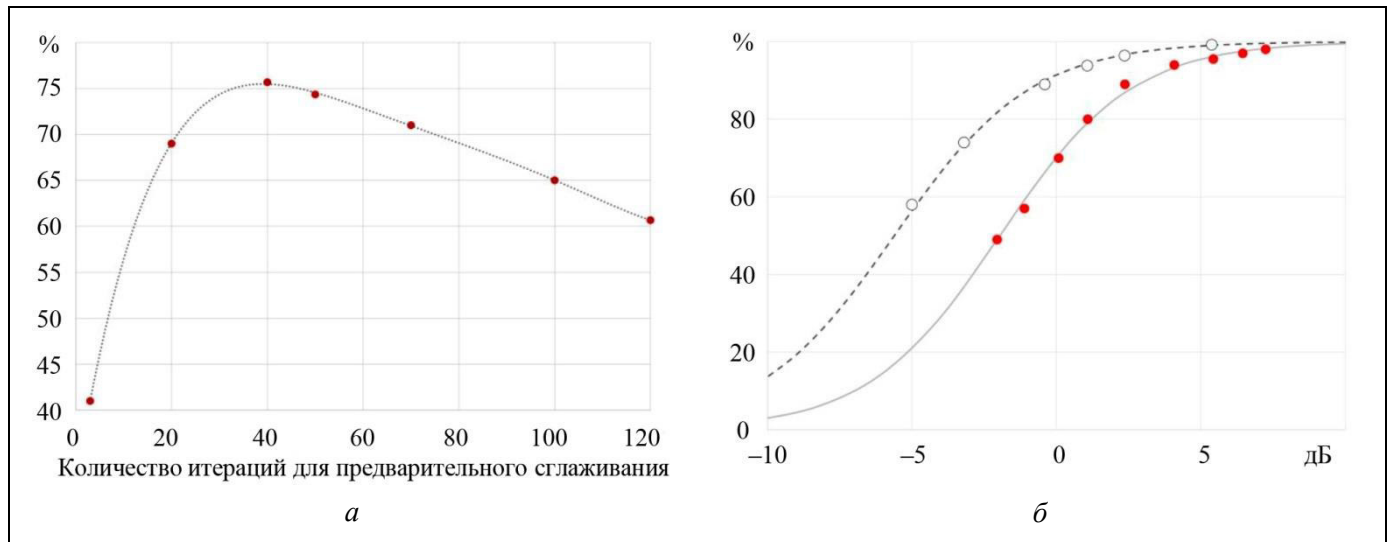


Рис. 4: a – зависимость доли безошибочных оценок ОБЛР Ω ($\text{SNR} = -3$ дБ) от предварительного сглаживания псевдоспектра гармоник; b – сравнение зависимостей безошибочности оценок ОБЛР при отсутствии сглаживания спектра и при сглаживании по 40 итерациям

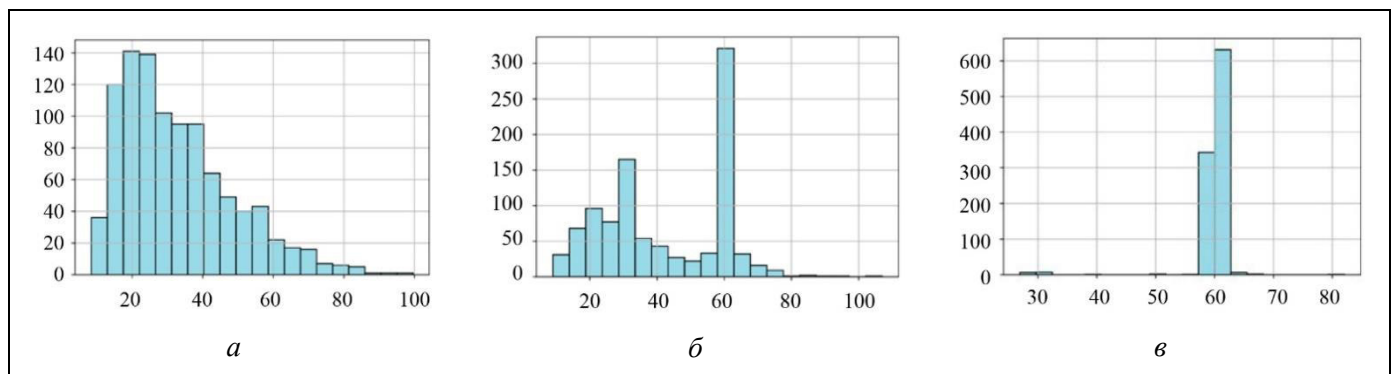


Рис. 5. Гистограммы ОБЛР для различных SNR (без сглаживания): a – $\text{SNR} = -12$ дБ, $\Omega = 1\%$; b – $\text{SNR} = -3$ дБ, $\Omega = 40\%$; c – $\text{SNR} = +7$ дБ, $\Omega = 98\%$

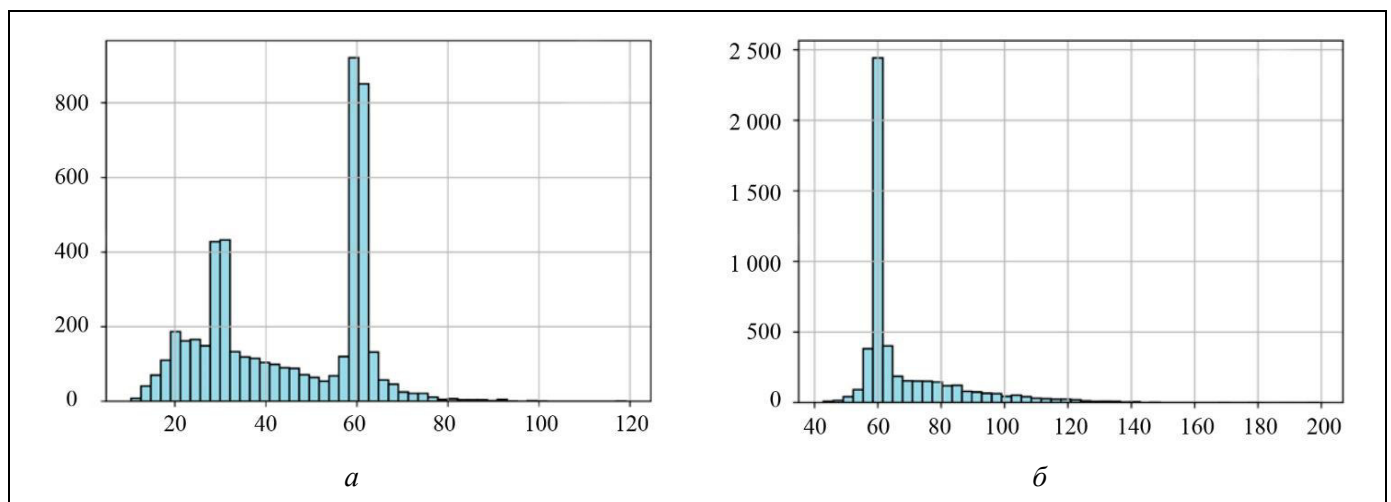


Рис. 6. Гистограммы ОБЛР для различных сглаживаний, $\text{SNR} = -3$ дБ: a – без сглаживания, $\Omega = 40\%$; b – со сглаживанием (100 итераций), $\Omega = 65\%$



Подведем итог и выпишем далее основные этапы алгоритма оценки основания гармонического ряда с помощью свертки:

1. Оценка периодограммы сигнала x_k длины N , размер временного окна T :

$$X_j = \sum_{k=0}^{N-1} x_k e^{-i \frac{2\pi}{N} jk}.$$

2. Логарифмирование энергетического спектра сигнала:

$$S_j^{(0)} = \ln |X_j|^2.$$

3. Выделение линейчатого (центрированного) спектра сигнала посредством свертки (M_0 – итерации размытия центрированного спектра, M – итерации сглаживания спектра для оценки огибающей, $M_0 \ll M$, $Tf_{\min} < j < Tf_{\max}$):

$$\hat{S}_k = S_j^{(0)} * \left[\frac{1}{\sqrt{M_0}} \exp\left(-\frac{j^2 T^2}{M_0}\right) - \frac{1}{\sqrt{M}} \exp\left(-\frac{j^2 T^2}{M}\right) \right],$$

из-за краевых эффектов среднее значение полученного ряда $\hat{S}_k$ немного отличается от нуля, и до построения автокорреляционной функции требуется из полученного ряда вычитать его среднее значение.

4. Оценка автокорреляционной функции центрированного спектра:

$$C_k^{(0)} = \sum_{j=Tf_{\min}}^{Tf_{\max}-k} \hat{S}_j \hat{S}_{j+k},$$

$$C_k^{(p)} = \sum_{j=Tf_{\min}}^{Tf_{\max}-k} C_j^{(p-1)} C_{j+k}^{(p-1)},$$

где $p \in [1, P]$, $P+1$ – количество последовательных автокорреляций.

5. Оценка ОВЛР:

$$f_0 = 2\pi f_H \left[\arctg \left(\frac{\mathcal{H}\{C^{(P)}(f_H)\}}{C^{(P)}(f_H)} \right) \right]^{-1},$$

где $f_{\min} < f_H < f_{\max}$.

3. ОПТИМАЛЬНЫЕ ПАРАМЕТРЫ АЛГОРИТМА ОПРЕДЕЛЕНИЯ ОВЛР

Пусть имеется некоторый набор значений ОВЛР, полученных в результате обработки данных звукозаписей шумов морских судов посредством вышеописанного алгоритма:

$$F_0^{(r)} = \{f_0^{(r)}\}_{i=1}^{Z_r},$$

где r – номер wav-файла; $F_0^{(r)}$ – набор значений оценок ОВЛР в каждом временном окне для wav-файла с номером r ; Z_r – количество временных окон в звуковом файле с номером r ; $Z_r T$ – размер звуковой записи морского судна.

Среди всех звуковых файлов отбираются только те, для которых справедливо неравенство $\sigma^{(r)} < \alpha T^{-1}$ ($\alpha \approx 10$), где $\sigma^{(r)}$ – стандартное отклонение оценок ОВЛР для записи r , т.е. в которых изменение ОВЛР за время наблюдения незначительно. Ошибочными считаются такие оценки ОВЛР, когда выполняется следующее условие:

$$|f_0^{(r)} - \mu^{(r)}| > \alpha T^{-1},$$

где $\mu^{(r)}$ – среднее значение оснований вально-лопастного ряда для записи r .

Критерий для поиска оптимального кортежа параметров $(T, M, f_{\min}, f_{\max}, P, f_H)$ алгоритма оценки основания вально-лопастного ряда определяется следующим образом:

$$\sum_r \omega^{(r)} \xrightarrow{T, M, f_{\min}, f_{\max}, P, f_H} \min,$$

где $\omega^{(r)}$ – количество ошибок для звукового wav-файла r .

Согласно этому критерию на реальных данных по грузовым и пассажирским морским судам были получены следующие оптимальные параметры алгоритма:

- $T = 10$ с – размер временного окна;
- количество итераций сглаживания $M = 100$ (предварительное сглаживание для размытия псевдоспектра $M_0 = 10$);
- $(f_{\min} = 2$ Гц; $f_{\max} = 200$ Гц) – интервал автокорреляции;
- $P = 2$ (количество последовательных автокорреляций $P+1 = 3$);
- частота для оценки «кепстральной фазы» $f_H = 25$ Гц.

ЗАКЛЮЧЕНИЕ

В настоящей работе представлены новые подходы к оценке фундаментальной частоты ряда гармоник по одному временному окну в условиях сильного шума. Алгоритм определения основания вально-лопастного ряда показал устойчивую работу в условиях, когда отношение сигнал/шум превышает -5 дБ. В дальнейших работах планируется усовершенствование метода для получения

уточненной оценки при непрерывной обработке сигнала по нескольким временным окнам, а также с использованием информационных характеристик [15]. В статье было показано, что эффективный результат может быть получен с помощью комплекса мер и верно подобранных значений параметров алгоритма.

В процессе исследования были проведены численные эксперименты на некотором количестве записей шумов морских судов (всего было использовано более 400 звукозаписей судов разного типа: пассажирские, контейнеровозы, танкеры, тягачи). Результаты численных экспериментов подтвердили эффективность представленных методов.

В ряде случаев было замечено, что когда источником шума морского судна являются несколько винтов, то наблюдаются биения. Вследствие чего по отдельным временным окнам сложно определить частоту оборотов вала (в этих окнах сигналы на вальных частотах приходили в противофазе). Здесь, в частности, приходится сталкиваться с проблемой определения достоверности оценки основания вально-лопастного ряда, что также является целью последующих исследований.

Кроме того, стоит сказать, что настоящая работа является также составной частью группы подходов по исследованию спектров шумов морских судов. Отметим, что выявление и использование дополнительных информационных критериев в других диапазонах спектра шума морского судна также значительно повышает качество оценки ОВЛР.

ЛИТЕРАТУРА

1. Урик Р.Дж. Основы гидроакустики. – Л.: Судостроение, 1978. [Urik, R.Dzh. Osnovy gidroakustiki. – Leningrad: Sudostroenie, 1978. (In Russian)]
2. Евтютов А.П., Митько В.Б. Примеры инженерных расчетов в гидроакустике. – Л.: Судостроение, 1981. [Evtyutov, A.P., Mit'ko, V.B. Primery inzhenernykh raschetov v gidroakustike. – Leningrad: Sudostroenie, 1981. (In Russian)]
3. Кудрявцев А.А., Лугинец К.П., Машошин А.И. Об амплитудной модуляции подводного шумоизлучения гражданских судов // Акустический журнал. – 2003. – № 49(2). – С. 224–228. [Kudryavcev, A.A., Lugines, K.P., Mashoshin, A.I. Ob amplitudnoj modulyacii podvodnogo shumozlucheniya grazhdanskih sudov // Akusticheskij zhurnal. – 2003. – No. 49 (2). – P. 224–228. (In Russian)]
4. Патент RU 2464588 C1. Устройство обнаружения шумовых гидроакустических сигналов в виде звукояда на основе вычисления интегрального вейвлет-спектра : 2011126056/28 : заявл. 15.06.2011 : опубл. 20.10.2012, бюл. № 29 // Малый В.В., Сапрыкин В.А., Рохманийко А.Ю. и др. [Patent RU 2464588 C1. Ustrojstvo obnaruzheniya шумовых gidroakusticheskikh signalov v vide zvukoryada na osnove vychisleniya integral'nogo vejvlet-spektra : 2011126056/28 : appl. 15.06.2011 : publ. 20.10.2012, byul. no. 29 / Malyj, V.V., Saprykin, V.A., Rohmanijko, A.Yu., et al. (In Russian)]
5. Бурдик В.С. Анализ гидроакустических систем. – Л.: Судостроение, 1988. – 364 с. [Burdik, V.S. Analiz gidroakusticheskikh sistem. – Leningrad: Sudostroenie, 1988. – 364 s. (In Russian)]
6. Евтютов А.П., Колесников А.Е., Корепин Е.А. и др. Справочник по гидроакустике. – Л.: Судостроение, 1988. [Evtyutov, A.P., Kolesnikov, A.E., Korepin, E.A. et al. Spravochnik po gidroakustike. – Leningrad: Sudostroenie, 1988. (In Russian)]
7. Astfalck, L.C., Sykulski, A.M., and Cripps, E.J. Debiasing Welch's Method for Spectral Density Estimation. – arXiv:2312.13643, 2023. – P. 1–17. – DOI: <https://doi.org/10.48550/arXiv.2312.13643>
8. Книхута Е.В., Пахотин В.А., Будник С.С., Ржанов А.А. Решение задачи оценки параметров сигнала в частотном пространстве // Известия вузов. Радиоэлектроника. – 2005. – № 2. – С. 19–29. [Knihuta, E.V., Pahotin, V.A., Budnik, S.S., Rzhanov, A.A. Reshenie zadachi ocenki parametrov signala v chastotnom prostranstve // Izvestiya vuzov. Radioelektronika. – 2005. – No. 2. – P. 19–29. (In Russian)]
9. Маранулец Ю.В. Адаптивный Спектральный анализ амплитудной огибающей шумов морских судов // Вестник Камчатского государственного технического университета. – 2003. – № 2. – С. 52–60. [Maranulec, Yu.V. Adaptivnyj Spektral'nyj analiz amplitudnoj ogibayushchej шумов morskikh sudov // Vestnik Kamchatskogo gosudarstvennogo tekhnicheskogo universiteta. – 2003. – No. 2. – P. 52–60. (In Russian)]
10. Liu, D., Yang, H., Hou, W., and Wang, B. A Novel Underwater Acoustic Target Recognition Method Based on MFCC and RACNN // Sensors. – 2024. – Vol. 24, no. 1. – Art. no. 273.
11. Doan, V.-S., Huynh-The, T., and Kim, D.-S. Underwater Acoustic Target Classification Based on Dense Convolutional Neural Network // IEEE Geoscience and Remote Sensing Letters. – 2022. – Vol. 19. – Art. no. 15009052020. – P. 1–5. – DOI: 10.1109/LGRS.2020.3029584
12. Randall, R.B. Frequency Analysis. – Naerum: Brüel & Kjaer, 1987. (In Russian)
13. Гольдин А.С. Вибрация роторных машин. – М.: Машиностроение, 2000. [Gol'din, A.S. Vibraciya rotornykh mashin. – M.: Mashinostroenie, 2000. (In Russian)]
14. Генкин М.Д., Соколова А.Г. Виброакустическая диагностика машин и механизмов. – М.: Машиностроение, 1987. [Genkin, M.D., Sokolova, A.G. Vibroakusticheskaya diagnostika mashin i mekhanizmov. – M.: Mashinostroenie, 1987. (In Russian)]
15. Галаев А.А., Бабилов В.Г., Лысенко П.В., Берлин Л.М. Новая спектральная мера сложности и ее возможности по обнаружению сигналов в шум // Доклады Российской академии наук. Математика, информатика, процессы управления. – 2024. – Т. 518. – С. 80–88. [Galyaev, A.A., Babikov, V.G., Lysenko, P.V., Berlin, L.M. A New Spectral Measure of Complexity and Its Capabilities for Detecting Signals in Noise // Doklady Mathematics. – 2024. – Vol. 110. – P. 361–368.]

Статья представлена к публикации членом редколлегии
Б. В. Павловым.

Поступила в редакцию 15.04.2025,
после доработки 05.06.2025.
Принята к публикации 25.06.2025.



Баби́ков Оле́г Влади́мирович – математик,
✉ babikov.ov@phystech.edu

© 2025 г. Баби́ков О. В., Баби́ков В. Г.

Баби́ков Влади́мир Гео́ргиевич – канд. физ.-мат. наук,
✉ babikov@ipu.ru



Институт проблем управления им. В.А. Трапезникова РАН,
г. Москва

Эта статья доступна по лицензии Creative Commons
«Attribution» («Атрибуция») 4.0 Всемирная.

ESTIMATING THE FUNDAMENTAL FREQUENCY OF A PROPELLER SHAFT-BLADE HARMONIC SERIES USING THE HILBERT TRANSFORM AND AUTOCORRELATION

O. V. Babikov* and V. G. Babikov**

***Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

*✉ babikov.ov@phystech.edu, **✉ babikov@ipu.ru

Abstract. The problem of estimating the fundamental frequency of a harmonic series arises in many areas of science and technology. For example, in vibration diagnosis, it is required to estimate the wear of bearings by the shift of the base of a harmonic series. In audio signal processing, this problem is associated with automatic instrument tuning. In speech synthesis, the fundamental frequency determines the pitch. In speech recognition, the frequency of the fundamental tone is an important information feature. In radio engineering, this problem is solved for the purpose of signal restoration, filtering, and decoding. In biomedical engineering, when analyzing a patient's ECG, EEG, voice, or breathing, pathologies such as arrhythmia are diagnosed by the fundamental frequency. In the detection and classification of sea vessels, the most significant information criterion is the base of a propeller shaft-blade harmonic series. This paper proposes new approaches to estimating the fundamental frequency in high noise conditions. To reduce errors, the idea is to use the method of periodograms, filtering, autocorrelation, and the Hilbert transform. Note that in high noise conditions, the estimate of the fundamental frequency of a harmonic series is significantly improved by selecting optimal parameters: the size of the time window, filtering parameters, the spectrum interval for autocorrelation, and the number of autocorrelations.

Keywords: fast Fourier transform, discrete Fourier transform, autocorrelation, Hilbert transform, fundamental frequency.

Acknowledgments. This work was supported in part by the Russian Science Foundation, project no. 23-19-00134.

МОДЕЛИ УТОМЛЕНИЯ И ОТДЫХА ПРИ НАУЧЕНИИ.

Ч. 1. Расширение общей модели итеративного научения

Д. И. Гребенков*, А. А. Козлова**, Д. В. Лемтюжникова***, Д. А. Новиков****

*-****Институт проблем управления им. В.А. Трапезникова РАН, г. Москва

*✉ grebenkov-d-i@mail.ru, **✉ sankamoro@mail.ru, ***✉ darabbt@gmail.com, ****✉ novikov@ipu.ru

Аннотация. Производится расширение классических моделей итеративного научения таким образом, чтобы учитывались факторы утомления и отдыха. На основе анализа существующих подходов к моделированию утомления и отдыха из различных областей – образования, производства, спорта, медицины – обоснована необходимость дополнения моделей этими процессами. Предложены математические модели, описывающие динамику уровня научения в периоды отдыха и с учетом снижения эффективности освоения навыков из-за утомления. Рассмотрены десять моделей возрастающей сложности: от базовых, без учета утомления, до комплексных, включающих зависимость вероятностей освоения и забывания от времени, а также от фазы отдыха. Показано, что наличие перерывов оптимальной продолжительности позволяет повысить терминальный уровень научения. При этом в модели с утомлением, отдыхом и без забывания оптимальный момент начала отдыха не зависит от меняющейся со временем вероятности освоения и приходится на середину периода формирования опыта. Полученные модели предназначены для прогнозирования производительности, оптимизации образовательных программ, производственных процессов и тренировочных циклов. Работа подчеркивает необходимость учета биологических и когнитивных ограничений при проектировании адаптивных систем обучения.

Ключевые слова: опыт, итеративное научение, кривая научения, математическое моделирование, утомление, отдых, освоение, забывание.

ВВЕДЕНИЕ

Научение – процесс и результат приобретения индивидуального опыта [1] – лежит в основе адаптации живых и неживых систем к изменяющимся условиям. В контексте математического моделирования под научением понимается процесс, в ходе которого система (биологическая, техническая или абстрактно-логическая) оптимизирует свои действия для достижения заданной цели. Особый интерес представляет *итеративное научение* (ИН) – вид обучения, основанный на многократном повторении системой действий, проб и ошибок для достижения фиксированной цели при постоянных внешних условиях [2]. Этот процесс формирует основу для выработки навыков у человека, условных рефлексов у животных, а также алгоритмов

адаптации в робототехнике и искусственном интеллекте.

Математические модели ИН описывают последовательность значений уровня научения – так называемые *кривые научения* (КН) [3] – через системы уравнений, графиков или алгоритмов, выявляя универсальные закономерности. Например, КН может быть зависимостью вероятности освоения компонента деятельности от времени или числа повторений (итераций).

Основываясь на экспериментальных данных, можно сделать вывод, что для большинства систем – от человека до нейронных сетей – эти кривые имеют замедленно-асимптотический характер: скорость улучшения показателей со временем снижается, а сама кривая стремится к некоторому пределу [2]. Такое поведение КН часто аппроксими-



мируется экспоненциальными функциями, что подтверждает «насыщение» процесса обучения.

Несмотря на универсальность замедленно-асимптотических закономерностей, классические модели научения не учитывают такие факторы, как утомление и отдых [3]. Для понимания этих ограничений и обоснования расширения моделей необходимо проанализировать существующие подходы к моделированию научения в различных контекстах – от когнитивных процессов в образовании до моторных навыков в спорте и производстве.

1. ПОДХОДЫ К МОДЕЛИРОВАНИЮ НАУЧЕНИЯ И ОТДЫХА: ОБЗОР

Научение играет ключевую роль практически в любой человеческой деятельности (*деятельность* – активное взаимодействие человека с окружающей действительностью, в ходе которого человек выступает как *субъект*, целенаправленно воздействующий на *предмет* [4]): от образования до промышленного производства, от спорта до роботизированной хирургии – понимание закономерностей научения позволяет оптимизировать процессы и снижать риски. Современные исследования демонстрируют, что моделирование научения требует учета не только индивидуальных особенностей, но и специфики контекста, будь то когнитивная нагрузка в образовании или физическое утомление в спорте или (и) на производстве.

В **образовании** научение тесно связано с управлением когнитивными ресурсами. Исследование влияния когнитивной усталости на результаты стандартизированных тестов выявило четкую зависимость между временем суток и успеваемостью [5]. Анализ двух миллионов тестов датских школьников показал, что каждый час после начала занятий результаты ухудшаются на 0,9 % стандартного отклонения. Это снижение объясняется накоплением усталости, которая нарушает концентрацию и уменьшает способность к решению сложных задач. Однако введение 20–30-минутных перерывов не только компенсирует этот эффект, но и повышает результаты на 1,7 %, особенно для учащихся с низкой успеваемостью. Для детей, испытывающих трудности в обучении, авторы работы [5] рекомендуют небольшие перерывы по 5–10 минут, которые снижают когнитивную перегрузку и улучшают усвоение материала на 12 %. Эти данные подчеркивают, что образовательный процесс должен проектироваться с учетом биологических

ритмов. Интересно, что аналогичные закономерности наблюдаются в спорте и производстве, где короткие паузы помогают сохранить продуктивность [6].

В работе [7] рассматривается моделирование процесса изучения второго языка и показывается, что прогресс в языковом навыке – это не плавный рост, а серия резких переходов между устойчивыми состояниями. Например, оптимальное соотношение нового и знакомого материала, примерно 30 % к 70 %, запускает заметный прогресс в освоении, тогда как дисбаланс приводит к «застреванию» процесса освоения на «плато». Это явление хорошо известно педагогам: студенты, сталкивающиеся с чрезмерно сложными заданиями, часто теряют мотивацию, в то время как рутинные упражнения без элементов новизны также тормозят развитие [8]. Такие выводы согласуются с корпусными исследованиями, где распределение ошибок в речи изучающих язык соответствует степенным законам [9, 10].

В **профессиональной деятельности**, особенно в высокотехнологичных отраслях, научение часто сопряжено с проблемой утомления. Исследование когнитивных эффектов утомления у сотрудников инженеринговых компаний [11] выявило, что после восьми часов работы точность поиска информации в памяти падает на 6 %, а время переключения между задачами увеличивается на 120 мс. Более того, 22 % сотрудников неосознанно переходят на менее эффективные стратегии решения задач, экономя когнитивные ресурсы. Эти данные особенно актуальны для сфер, где ошибки могут иметь критические последствия, – от авиадиспетчерских служб до медицинской диагностики. В статье [11] рекомендуется внедрять адаптивные графики работы, где периоды интенсивной нагрузки чередуются с фазами отдыха, а также использовать тренажеры для отработки навыков в условиях, имитирующих утомление.

Особый интерес представляет освоение сложных медицинских технологий, таких как роботизированная хирургия. Исследование кривых обучения для медсестер, работающих в операционных [12], продемонстрировало, что достижение мастерства требует в среднем 8–11 процедур. Однако индивидуальные различия огромны: некоторым достаточно трех операций, другим – до 11-ти. Наиболее сложными этапами оказались планирование установки спинальных винтов и управление роботизированной рукой, где ошибки часто связаны с когнитивной перегрузкой. Для сокращения времени обучения авторы работы [12] предлагают

симуляционные тренировки на 3D-моделях позвоночника, которые позволяют отработать ключевые навыки без риска для пациентов. Этот подход снижает уровень стресса и утомления на 35 %, что напрямую влияет на скорость освоения технологии.

В **производственной сфере** моделирование научения стало основой для оптимизации процессов. Классическая степенная кривая Райта [13], описывающая снижение времени выполнения задач с накоплением опыта, до сих пор остается отправной точкой. Однако многие задачи, сочетающие когнитивные и моторные компоненты, требуют более сложных подходов. Модель, представленная в статье [14], разделяет время выполнения задачи на два компонента: когнитивный (планирование, принятие решений) и моторный (физические действия). Например, при сборке механического устройства начальные этапы требуют активного мышления, и их время сокращается быстрее, тогда как моторные навыки улучшаются постепенно. Эта модель на 23 % точнее предсказывает производительность на ранних стадиях обучения по сравнению с традиционными подходами, что критически важно для расчета себестоимости и планирования производства.

В метаанализе [15], охватившем 115 наборов данных, систематизированы лучшие практики для моделирования научения в производстве. Для задач, связанных со снижением времени или затрат, наиболее эффективными оказались S-образная и трехпараметрическая гиперболическая модели. Они учитывают фазу «плато», когда дальнейшее улучшение становится минимальным. В контексте роста производительности лидирует трехпараметрическая экспоненциальная модель, которая хорошо описывает процессы с эффектом насыщения. Интеграция этих моделей в системы управления производством сокращает ошибки в планировании запасов на 18 %, предотвращая как избыточные складские запасы, так и простои.

В **спорте и реабилитации** научение неотделимо от управления утомлением. Исследование [16] показало, что утомление увеличивает риск травм на 40 % из-за снижения проприоцепции – способности ощущать положение тела в пространстве. Например, в футболе 67 % травм случаются в последние 20 минут матча, когда концентрация внимания падает, а мышцы теряют эластичность. Для мониторинга состояния спортсменов используется комбинация объективных методов (анализ вариабельности сердечного ритма) и субъективных опросников. Анализ вариабельности сердечного

ритма позволяет прогнозировать готовность к нагрузке: снижение вариабельности на 15 % коррелирует с повышением риска травм. Эти данные используются для персонализации тренировок – сокращения интенсивности при первых признаках переутомления.

Отдельного внимания заслуживает проблема реабилитации после травм или инсультов. В модели, предложенной в статье [17], утомление рассматривается как баланс между двумя компонентами: объективным снижением производительности и субъективным восприятием усилий. Например, пациенты, перенесшие инсульт, часто ощущают слабость мышц даже при минимальной нагрузке. Традиционные тренировки, направленные на увеличение силы, могут усугубить состояние, если не учитывают индивидуальные пороги утомления. Авторы работы [17] предлагают использовать сенсорную обратную связь (например, визуализацию мышечной активности в реальном времени), которая помогает пациентам учиться распределять усилия, избегая перенапряжения.

В некоторых работах по **роботизированным системам и промышленному дизайну** процесс научения описывается с помощью биомеханического анализа. Например, в статье [6] представлены дифференциальные уравнения, описывающие динамику мышечного утомления. Восстановление после нагрузки мышцы в 5 кг требует в среднем 2,4 минуты отдыха – эти данные используются для проектирования рабочих мест на конвейерах. Симуляторы, интегрирующие такие модели, позволяют тестировать эргономику еще до запуска производства. В авиационной промышленности это сократило число случаев кумулятивного утомления на 22 % благодаря оптимизации углов наклона инструментов и высоты рабочих поверхностей.

Таким образом, современные исследования процесса научения демонстрируют, что универсальных моделей научения, учитывающих утомление, не существует – каждый контекст требует учета своей специфики. В образовании акцент в настоящее время смещается на управление когнитивной нагрузкой и временными промежутками, в производстве – на разделение когнитивных и моторных компонентов, в спорте – на баланс между утомлением и восстановлением. Как показывают рассмотренные подходы, утомление и отдых играют важную роль в динамике формирования опыта. Следовательно, необходимо дополнить существующие классические модели научения учетом этих процессов. Возьмем в качестве основы для расширения модели из работы [3].



2. УТОМЛЕНИЕ И ОТДЫХ В МОДЕЛЯХ ФОРМИРОВАНИЯ ОПЫТА

Рассмотрим следующие модели научения. Пусть обучаемый (агент) приобретает индивидуальный опыт – осваивает некоторый вид деятельности посредством последовательных попыток (т. е. через ИН), причем опыт в каждом периоде характеризуется двумя возможными состояниями – «сформирован» или «не сформирован». В каждый период времени $t = 1, 2, \dots$, если опыт не сформирован, происходит формирование опыта с вероятностью $0 \leq w(t) < 1$, зависящей в общем случае от времени. Параллельно происходит забывание опыта с вероятностью $0 \leq u(t) < 1$, также зависящей в общем случае от времени. Значение $q(t)$ критерия индивидуального опыта агента принадлежит единичному отрезку. Значением $q(t)$ будем считать вероятность того, что в периоде t опыт агента будет сформирован и не забыт. Как зависимость вероятности формирования опыта от времени (убывающая), так и зависимость от него (возрастающая) вероятности забывания могут отражать эффекты утомления, усталости и пр. в процессе учебной и (или) продуктивной деятельности.

Рассмотрим последовательно (в порядке усложнения) ряд моделей научения, учитывающих такие процессы, как освоение, забывание, утомление и отдых. Наличие тех или иных процессов в моделях отражено в таблице.

Модель 1. Базовой (простейшей) является модель научения [3], в которой забывание отсутствует, а вероятность освоения опыта w не зависит от времени. В рамках этой модели динамика уровня научения имеет вид:

$$q(t+1) = q(t) + (1 - q(t))w, \quad t = 0, 1, 2, \dots, \quad (1)$$

где начальное значение $q(0)$ считается известным.

В непрерывном времени разностному уравнению (1) соответствует дифференциальное уравнение

$$\dot{q} = (1 - q)w,$$

имеющее решение

$$q(t) = 1 - (1 - q(0))e^{-wt}. \quad (2)$$

Кривая научения (2) не убывает и асимптотически стремится к единице. График функции (2) при $q(0) = 0$ продемонстрирован на рис. 1.

Процессы и параметры моделей формирования опыта

Модели	Освоение	Забывание	Утомление	Отдых
Модель 1	Не зависит от времени	Нет	Нет	Нет
Модель 2	Нет	Не зависит от времени	Нет	Нет
Модель 3	Не зависит от времени	Не зависит от времени	Нет	Нет
Модель 4	Зависит от времени	Нет	Есть	Нет
Модель 5	Зависит от времени	Не зависит от времени	Есть	Нет
Модель 6	Не зависит от времени	Зависит от времени	Есть	Нет
Модель 7	Зависит от времени	Нет	Есть	Есть
Модель 8	Зависит от времени	Не зависит от времени, учитывается только во время отдыха	Есть	Есть
Модель 9	Зависит от времени	Не зависит от времени	Есть	Есть
Модель 10	Не зависит от времени	Зависит от времени	Есть	Есть

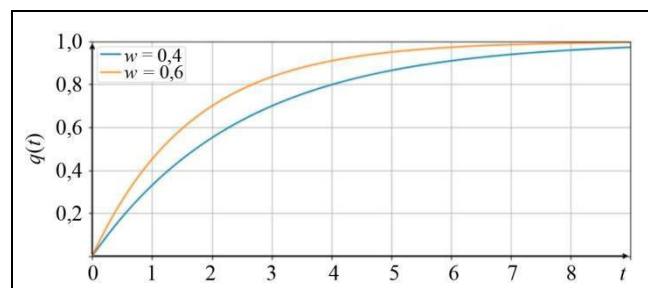


Рис. 1. Пример кривой научения для модели 1

Модель 2 (модель забывания). Предположим, что освоения опыта не происходит, а имеет место только забывание, не зависящее от времени.

Получим:

$$\begin{aligned} q(t+1) &= q(t) - uq(t), \\ \dot{q} &= -qu, \\ q(t) &= q(0)e^{-ut}. \end{aligned} \quad (3)$$

График функции (3) при $q(0)=1$ приведен на рис. 2.

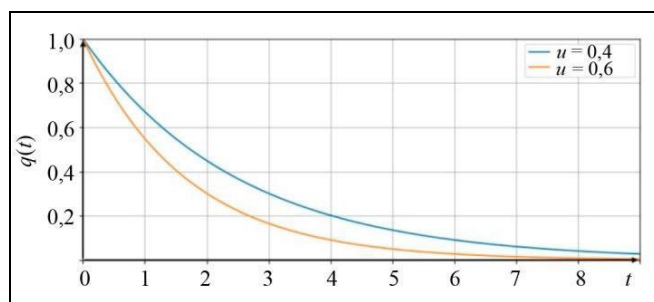


Рис. 2. Пример кривой научения для модели 2

Модель 3. Пусть в условиях модели 1 дополнительно имеет место забывание (или в рамках модели 2 имеет место научение), не зависящее от времени. Получим:

$$\begin{aligned} q(t+1) &= q(t) + (1-q(t))w - uq(t), \\ \dot{q} &= (1-q)w - qu, \end{aligned} \quad (4)$$

$$q(t) = \frac{w}{w+u} - \left(\frac{w}{w+u} - q(0) \right) e^{-(w+u)t}. \quad (5)$$

Кривая научения (5) при $q(0) < \frac{w}{w+u}$ не убывает и асимптотически стремится к $\frac{w}{w+u}$. График функции (5) при $q(0)=0$ и $w=0,6$ приведен на рис. 3.

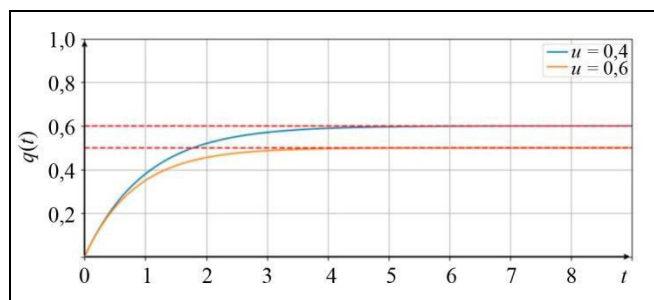


Рис. 3. Пример кривой научения для модели 3

Модель 4. Пусть в условиях модели 1 вероятность формирования опыта $w(t) > 0$ зависит от

времени. Предположим, что $w(\cdot)$ – непрерывная невозрастающая функция (ее можно условно назвать «кривой утомления»), что содержательно может отражать утомление и (или) усталость агента в процессе научения. Получим:

$$\begin{aligned} q(t+1) &= q(t) + w(t)(1-q(t)), \\ \dot{q} &= (1-q)w(t), \\ q(t) &= 1 - (1-q(0))e^{-W(t)}, \end{aligned} \quad (6)$$

где $W(t) = \int_0^t w(\xi) d\xi$. Очевидно, что эта функция обладает следующими свойствами.

Лемма 1. $W(\cdot)$ – непрерывная положительно-значная монотонно возрастающая вогнутая функция времени, $W(0) = 0$.

Кривая научения (6) не убывает и асимптотически стремится к единице. В частном случае, при $w(t) = w$, модель 3 переходит в модель 1.

Для иллюстраций в качестве $w(t)$ здесь и далее будем использовать $w(t) = w_0 e^{-\alpha t}$ и $W(t) = \frac{w_0}{\alpha} (1 - e^{-\alpha t})$. График функции (6) при $q(0)=0$ и $w_0=0,6$ приведен на рис. 4.

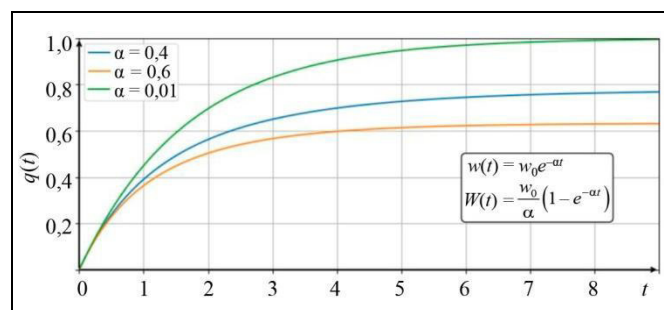


Рис. 4. Пример кривой научения для модели 4

Модель 5. Пусть в условиях модели 4 имеет место постоянное во времени забывание $u(t) = u$. Получим:

$$\begin{aligned} q(t+1) &= q(t) + w(t)(1-q(t)) - uq(t), \\ \dot{q} &= (1-q)w(t) - uq, \\ q(t) &= e^{-(W(t)+ut)} \cdot \int_0^t w(\xi) \cdot e^{(W(\xi)+u\xi)} d\xi. \end{aligned} \quad (7)$$

График функции (7) при $w_0=0,6$ и $u=0,1$ приведен на рис. 5. Видно, что кривая научения в модели 5 имеет точку максимума.

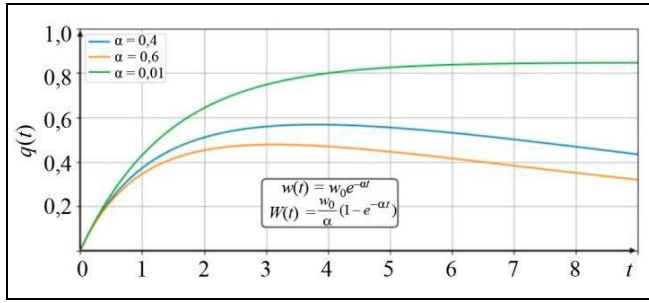


Рис. 5. Пример кривой обучения для модели 5

Модель 6. Пусть в условиях модели 3 вероятность забывания зависит от времени: $u(\cdot)$ – непрерывная неубывающая функция. Тогда решением уравнения (4) будет кривая обучения

$$q(t) = e^{-U(t)} \left(q(0) + w \int_0^t e^{U(\xi)} d\xi \right), \quad (8)$$

где $U(t) = \int_0^t (w + u(\xi)) d\xi$. При $U(t) = u$ выражение (8) переходит в выражение (5). График функции (8) при $w = 0,6$, $q(0) = 0$, $u(t) = 1 - e^{-at}$ и $U(t) = (w+1)t - \frac{1}{a}(1 - e^{-at})$ приведен на рис. 6.

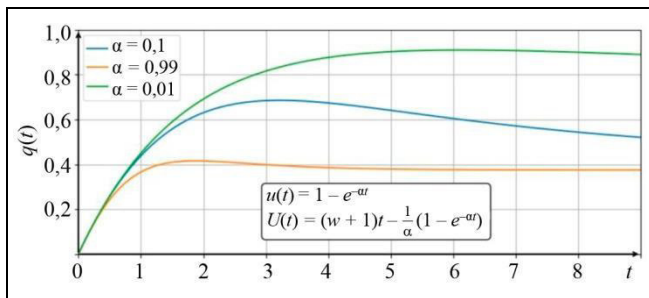


Рис. 6. Пример кривой обучения для модели 6

Лемма 2. $U(\cdot)$ – непрерывная положительно-значная монотонно возрастающая выпуклая функция, $U(0) = 0$.

График функции (8) при $w_0 = 0,6$ и $u_0 = 0,1$ приведен на рис. 5.

Кривая (8) может иметь точку максимума при значении t^* , удовлетворяющем условию

$$(w + u(t^*)) \int_0^{t^*} e^{U(\xi)} d\xi = e^{U(t^*)}.$$

Содержательная интерпретация t^* – максимальная целесообразная продолжительность обучения.

Модель 7. Пусть в условиях модели 4 агент отдыхает в промежутке времени $[\tau; \tau + \Delta]$, где $\tau > 0$ – момент начала отдыха, $\Delta \geq 0$ – его продолжительность. Так как забывание в рассматриваемом случае отсутствует, предположим, что $q(\tau + \Delta) = q(\tau)$, а результат отдыха с точки зрения роста вероятности формирования опыта здесь и далее отразим предположением, что продолжительности отдыха $\Delta \geq \Delta_0$ достаточно для полного восстановления агента (где минимальная необходимая продолжительность отдыха $\Delta_0 > 0$ известна), т. е.

$$w(t) = w(t - \tau - \Delta), \quad t \geq \tau + \Delta.$$

Тогда

$$q(t) = \begin{cases} 1 - (1 - q(0))e^{-W(t)}, & t \in [0; \tau]; \\ 1 - (1 - q(0))e^{-W(\tau)}, & t \in [\tau; \tau + \Delta]; \\ 1 - [(1 - q(0))e^{-W(\tau)}]e^{-W(t - \tau - \Delta)}, & t \geq \tau + \Delta. \end{cases} \quad (9)$$

Далее для простоты будем считать начальный уровень обучения равным нулю: $q(0) = 0$. График функции (9) при $w_0 = 0,6$ и $\alpha = 0,6$ приведён на рис. 7.

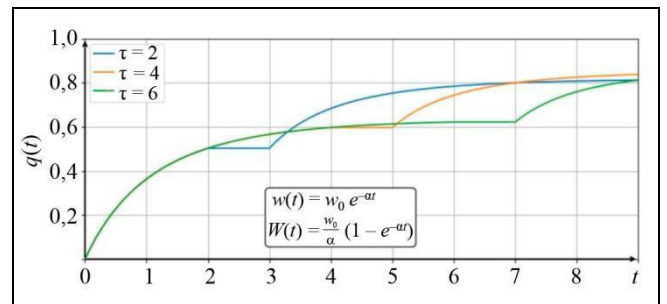


Рис. 7. Пример кривой обучения для модели 7

Выбирая время и продолжительность отдыха, можно ставить и решать две задачи.

Первая задача – максимизация терминального (в момент времени $T > 0$) значения уровня обучения, называемого горизонтом планирования:

$$1 - e^{-W(\tau) - W(T - \tau - \Delta)} \rightarrow \max_{\tau \in (0; T - \Delta_0), \Delta \in [\Delta_0; T - \tau]}. \quad (10)$$

Вторая задача – максимизация интегрального значения уровня обучения:

$$\begin{aligned} \int_0^{\tau} (1 - e^{-W(\xi)}) d\xi + \int_{\tau+\Delta}^T (1 - e^{-W(\tau)} e^{-W(\xi - \tau - \Delta)}) d\xi \rightarrow \\ \rightarrow \max_{\tau \in (0; T - \Delta_0), \Delta \in [\Delta_0; T - \tau]}. \end{aligned} \quad (11)$$

Содержательная интерпретация интегрального критерия – объем правильно выполненных агентом действий, если научение происходит в процессе продуктивной деятельности [18].

Задачу (10) можно записать в виде:

$$W(\tau) + W(T - \tau - \Delta) \rightarrow \max_{\tau \in (0; T - \Delta_0), \Delta \in [\Delta_0; T - \tau]} \quad (12)$$

Задачу (11) можно записать в виде:

$$\int_0^{\tau} e^{-W(\zeta)} d\zeta + e^{-W(\tau)} \int_0^{T-\tau-\Delta} e^{-W(\xi)} d\xi \rightarrow \min_{\tau \in (0; T - \Delta_0), \Delta \in [\Delta_0; T - \tau]} \quad (13)$$

В силу леммы 1 и структуры критериев (12) и (13), справедливо следующее

Утверждение 1. В модели 7 оптимальная продолжительность отдыха Δ^* (и в задаче максимизации терминального уровня научения, и в задаче максимизации интегрального уровня научения) равна минимально необходимой, т. е. $\Delta^* = \Delta_0$.

С учетом результата утверждения 1 задачи (12) и (13) сводятся соответственно к следующим задачам выбора оптимального момента начала отдыха:

$$W(\tau) + W(T - \tau - \Delta_0) \rightarrow \max_{\tau \in (0; T - \Delta_0)} \quad (14)$$

$$\int_0^{\tau} e^{-W(\zeta)} d\zeta + e^{-W(\tau)} \int_0^{T-\tau-\Delta_0} e^{-W(\xi)} d\xi \rightarrow \min_{\tau \in (0; T - \Delta_0)} \quad (15)$$

Условие первого порядка для решения τ^* задачи (14) имеет вид: $w(\tau^*) = w(T - \tau^* - \Delta_0)$, откуда в силу непрерывности и монотонности функции $w(\cdot)$ получаем:

$$\tau^* = \frac{T - \Delta_0}{2} \quad (16)$$

Отрицательность второй производной (14) в этой точке легко проверяется.

Таким образом, согласно модели 7, с точки зрения максимизации терминального значения уровня научения перерыв лучше совершать примерно в середине периода формирования опыта.

Утверждение 2. В модели 7 оптимальный (с точки зрения максимизации терминального уровня научения) момент начала отдыха определяется выражением (16) и не зависит от кривой утомления.

Выражение (16) дает оптимальный момент начала отдыха для случая, если отдых продолжи-

тельностью Δ_0 обязательно должен быть осуществлен. Для ответа на вопрос, необходим ли перерыв на отдых, следует сравнить значения критерия (10) в точке τ^* с максимальным значением уровня научения $1 - e^{-W(T)}$, которое может быть достигнуто при отсутствии отдыха. Ответ на этот вопрос, конечно, зависит и от свойств функции $w(\cdot)$, и от значений величин T и Δ_0 , и дается следующим утверждением.

Утверждение 3. Перерыв на отдых целесообразен с точки зрения максимизации терминального уровня научения, если выполнено

$$W\left(\frac{T - \Delta_0}{2}\right) \geq \frac{W(T)}{2} \quad (17)$$

Условие (17) можно записать в виде:

$$\int_0^{\frac{T - \Delta_0}{2}} w(\zeta) d\zeta \geq \int_{\frac{T - \Delta_0}{2}}^T w(\zeta) d\zeta \quad (18)$$

В силу убывания подынтегральной функции при достаточно малых Δ_0 и (или) достаточно больших T , условие (18) выполнено. В то же время при фиксированном Δ_0 можно найти значение T , обращающее выражение (18) в равенство – минимальный горизонт планирования, при котором отдых данной продолжительности будет еще целесообразен. И наоборот, при фиксированном T можно найти значение Δ_0 , обращающее выражение (18) в равенство – максимальную продолжительность отдыха, еще целесообразную при заданном горизонте планирования.

Перейдем к задаче (15). Условие первого порядка для ее решения τ^{**} имеет вид:

$$\int_0^{T-\tau^{**}-\Delta_0} e^{-W(\xi)} d\xi = \frac{1 - e^{-W(T-\tau^{**}-\Delta_0)}}{w(\tau^{**})}$$

Отрицательность второй производной (15) в этой точке легко проверяется.

Модель 8. Пусть в условиях модели 7 во время отдыха (и только во время него) имеет место забывание с постоянной и не зависящей от времени вероятностью u . Тогда в соответствии с выражением (3) модели 2 получим, что к моменту окончания отдыха уровень научения снизится до

$$q(t + \Delta_0) = q(\tau) e^{-u\Delta_0}$$



Тогда

$$q(t) = \begin{cases} 1 - e^{-W(t)}, & t \in [0; \tau]; \\ \left(1 - e^{-W(\tau)}\right) e^{-u(t-\tau)}, & t \in [\tau; \tau + \Delta_0]; \\ 1 - \left[1 - \left(1 - e^{-W(\tau)}\right) e^{-u \Delta_0}\right] e^{-W(t-\tau-\Delta_0)}, & t \geq \tau + \Delta_0. \end{cases} \quad (19)$$

Рассмотрим для модели 8 задачу максимизации терминального значения уровня научения. Интегральный критерий исследуется аналогично тому, как это делалось в модели 7.

$$1 - \left[1 - \left(1 - e^{-W(\tau)}\right) e^{-u \Delta_0}\right] e^{-W(t-\tau-\Delta_0)} \rightarrow \max_{\tau \in (0; T-\Delta_0)},$$

т. е.

$$\left[1 - \left(1 - e^{-W(\tau)}\right) e^{-u \Delta_0}\right] e^{-W(T-\tau-\Delta_0)} \rightarrow \min_{\tau \in (0; T-\Delta_0)}. \quad (20)$$

Условие первого порядка для задачи (20) имеет вид:

$$e^{u \Delta_0} - 1 = \left(\frac{w(\tau^*)}{w(T - \tau^* - \Delta_0)} - 1 \right) e^{-W(\tau^*)}.$$

График функции (19) при $w_0 = 0,6$, $u_0 = 0,6$ и $\alpha = 0,6$ приведен на рис. 8.

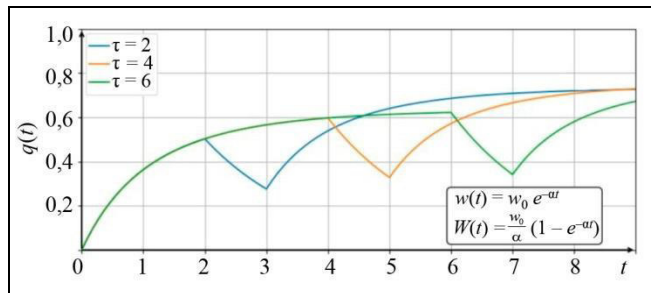


Рис. 8. Пример кривой научения для модели 8

Модель 9. Пусть в условиях модели 8 постоянное во времени забывание имеет место не только во время отдыха, но и во время научения. Участок $q(t)$ до отдыха в этом случае будет аналогичен формуле (7) из модели 5. Соответствующая кривая научения может уже не быть монотонной и может, например, иметь точку максимума.

Модель 10. Пусть в условиях модели 9 освоение не зависит от времени, а забывание — зависит. Участок $q(t)$ до отдыха в этом случае будет соответствовать формуле (8) из модели 6. Тогда аналог

выражения (19) для этой модели будет выглядеть так:

$$q(t) = \begin{cases} e^{-U(t)} \cdot \left(w \int_0^t e^{U(\xi)} d\xi + q(0) \right), & t \in [0; \tau]; \\ e^{\hat{U}(t-\tau)} \cdot \left[e^{-U(\tau)} \cdot \left(w \int_0^\tau e^{U(\xi)} d\xi + q(0) \right) \right], & t \in [\tau; \tau + \Delta_0]; \\ e^{-U(t-\tau-\Delta_0)} \cdot \left\{ w \int_0^{t-\tau-\Delta_0} e^{U(\xi)} d\xi + e^{\hat{U}(t-\tau)} \cdot \left[e^{-U(\tau)} \cdot \left(w \int_0^\tau e^{U(\xi)} d\xi + q(0) \right) \right] \right\}, & t \geq \tau + \Delta_0, \end{cases}$$

где в общем случае $\hat{U}(t) \neq U(t)$.

Дальнейшее аналитическое исследование этой и более сложных моделей не представляется целесообразным, так как полученные результаты громоздки и не дают сделать аналитических выводов о закономерностях формирования опыта. В последующих расчетах возможно применение численных методов. Однако для этого необходимо использовать конкретные зависимости вероятностей освоения и забывания от времени.

ЗАКЛЮЧЕНИЕ

В настоящей статье предложено расширение моделей формирования опыта из работы [3] с учетом факторов утомления и отдыха. Важным результатом стало выявление оптимальных временных интервалов для перерывов: для задач, которые описывает модель 7, с точки зрения максимизации терминального значения уровня научения оптимальный момент начала отдыха не зависит от меняющейся со временем вероятности освоения и находится примерно в середине периода формирования опыта. Кроме того, в более сложных моделях, зная функции освоения и (или) забывания, можно выявлять максимальное целесообразное время научения и оптимальный момент начала отдыха.

Таким образом, дополнение моделей формирования опыта учетом процессов утомления и отдыха повышает их практическую применимость, позволяя прогнозировать снижение эффективности и

оптимизировать ресурсы благодаря расчету интервалов отдыха.

Вторая часть статьи будет посвящена обзору наборов данных с динамикой формирования опыта, а также примерам применения приведенных выше моделей для описания результатов различных известных экспериментальных исследований.

ЛИТЕРАТУРА

1. *Краткий психологический словарь*. – М.: Политиздат, 1985. – 201 с. [*Kratkii psikhologicheskii slovar'*. – М.: Politizdat, 1985. – 201 s. (In Russian)]
2. *Новиков Д.А. Закономерности итеративного научения*. – М.: Институт проблем управления РАН, 1998. – 96 с. [*Novikov, D.A. Zakonomernosti iterativnogo naucheniya*. – М.: Institut problem upravleniya RAN, 1998. – 96 s. (In Russian)]
3. *Белов М.В., Новиков Д.А. Модели опыта // Проблемы управления*. – 2021. – № 1. – С. 43–60. – DOI: 10.25728/pu.2021.1.5. [*Belov, M.V., Novikov, D.A. Models of Experience // Control Sciences*. – 2021. – No. 1. – P. 37–52. – DOI: 10.25728/cs.2021.1.5. (In Russian)]
4. *Новиков А.М., Новиков Д.А. Методология*. – М.: Синтег, 2007. – 668 с. [*Novikov, A.M., Novikov, D.A. Metodologiya*. – М.: Sinteg, 2007. – 668 s. (In Russian)]
5. *Piovesan, M. Cognitive Fatigue Influences Students' Performance on Standardized Tests // Proceedings of the National Academy of Sciences of the United States of America*. – 2016. – Vol. 113, no. 10. – DOI: 10.1073/pnas.1516947113
6. *Bennis, F. A New Muscle Fatigue and Recovery Model and Its Ergonomics Application in Human Simulation // Virtual and Physical Prototyping*. – 2010. – Vol. 5, no. 3. – DOI: 10.1080/17452759.2010.504056
7. *van Geert, P. Some Thoughts on Dynamic Systems Modeling of L2 Learning // Frontiers in Physics*. – 2023. – Vol. 11. – DOI: 10.3389/fphy.2023.1186136
8. *Ryan, R.M., Deci, E.L. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being // American Psychologist*. – 2000. – Vol. 55, no. 1. – P. 68–78.
9. *Ellis, N., Larsen-Freeman, D. Constructing a Second Language: Analyses and Computational Simulations of the Emergence of Linguistic Constructions from Usage // Lang. Learn.* – 2009. – Vol. 59 (suppl. 1). – P. 90–125. – DOI: 10.1111/j.1467-9922.2009.00537.x
10. *Rastelli, S. The Discontinuity Model: Statistical and Grammatical Learning in Adult Second-Language Acquisition // Lang. Acquis.* – 2019. – Vol. 26, no. 4. – P. 387–415. – DOI: 10.1080/10489223.2019.1571594.
11. *Величковский Б.Б. Когнитивные эффекты умственного утомления // Вестник Московского университета. Сер. 14. Психология*. – 2019. – № 1. – С. 108–122. – DOI: 10.11621/vsp.2019.01.108. [*Velichkovsky, B.B. Cognitive Effects of Mental Fatigue // Lomonosov Psychology Journal*. – 2019. – No. 1. – P. 108–122. – DOI: 10.11621/vsp.2019.01.108. (In Russian)]
12. *Yao, Y., Wang, H., Zhang, Q., et al. Learning Curves for Itinerant Nurses to Master the Operation Skill of Ti-Robot Assisted Spinal Surgery Equipment by CUSUM Analysis: A Pilot Study // PLoS ONE*. – 2024. – Vol. 19, iss. 3. – DOI: 10.1371/journal.pone.0291147
13. *Wright, T.P. Factors Affecting the Cost of Airplanes // Journal of the Aeronautical Sciences*. – 1936. – Vol. 3, iss. 4. – P. 122–128.
14. *Mohamad, Y.J., Christoph, H.G. A Learning Curve for Tasks with Cognitive and Motor Elements // Computers & Industrial Engineering*. – 2013. – Vol. 64, iss. 3. – P. 866–871. – DOI: 10.1016/j.cie.2012.12.005
15. *Eric, H.G., Christoph, H.G., Sebastian, M. Production Economics and the Learning Curve: A Meta-analysis // International Journal of Production Economics*. – Vol. 170, part B. – P. 401–412. – DOI: 10.1016/j.ijpe.2015.06.021
16. *Bestwick-Stevenson, T., Toone, R., Neupert, E., et al. Assessment of Fatigue and Recovery in Sport: Narrative Review // International Journal of Sports Medicine*. – 2022. – Vol. 43, no. 14 – DOI: 10.1055/a-1834-7177
17. *Enoka, R.M., Duchateau, J. Translating Fatigue to Human Performance // Medicine & Science in Sports & Exercise*. – 2016. – Vol. 48, iss. 11. – P. 2228–2238. – DOI: 10.1249/MSS.0000000000000929.
18. *Белов М.В., Новиков Д.А. Модели технологий*. – М.: Ленанд, 2019. – 160 с. [*Belov, M.V., Novikov, D.A. Modeli tekhnologii*. – М.: Lenand, 2019. – 160 s. (In Russian)]

Статья представлена к публикации руководителем РРС
А. А. Ворониным.

Поступила в редакцию 26.03.2025,
после доработки 29.04.2025.
Принята к публикации 05.05.2025.

Гребенков Дмитрий Игоревич – математик,
✉ grebenkov-d-i@mail.ru,
ORCID iD: <https://orcid.org/0009-0002-7085-5912>

Козлова Анастасия Андреевна – инженер,
✉ sankamoro@mail.ru,
ORCID iD: <https://orcid.org/0009-0005-6105-121X>

Лемтюжникова Дарьяна Владимировна – канд. физ.-мат. наук, ✉ darabtt@gmail.com,
ORCID iD: <https://orcid.org/0000-0002-5311-5552>

Новиков Дмитрий Александрович – д-р техн. наук, академик РАН, ✉ novikov@ipu.ru,
ORCID iD: <https://orcid.org/0000-0002-9314-3304>

Институт проблем управления им. В.А. Трапезникова РАН,
г. Москва.

© 2025 г. Гребенков Д. И., Козлова А. А., Лемтюжникова Д. В., Новиков Д. А.



Эта статья доступна по лицензии Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная.



MODELS OF FATIGUE AND REST IN LEARNING. PART I: Extension of the General Iterative Learning Model

D. I. Grebenkov*, A. A. Kozlova**, D. V. Lemtyuzhnikova***, and D. A. Novikov****

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

*✉ grebenkov-d-i@mail.ru, **✉ sankamoro@mail.ru, ***✉ darabbt@gmail.com, ****✉ novikov@ipu.ru

Abstract. In this paper, classical iterative learning models are extended by including the factors of fatigue and rest. The existing approaches to model fatigue and rest from various fields—education, production, sports, and medicine—are analyzed, and the need to include these factors in the models is justified accordingly. Mathematical models are proposed to describe learning level dynamics during rest periods, considering the reduced efficiency of acquiring skills due to fatigue. Ten models of growing complexity are studied: from simple models without any fatigue effects to complex ones with the probabilities of skill acquisition and forgetting depending on time and rest periods. As is shown, the breaks of optimal duration allow improving the terminal learning level. In the model with fatigue, rest, and no forgetting, the optimal time to start rest is independent of the probability of skill acquisition as a function of time and lies at the middle of the experience acquisition interval. The models proposed are intended for predicting performance and optimizing training programs, production processes, and training cycles. This study emphasizes the need to consider biological and cognitive constraints when designing adaptive learning systems.

Keywords: experience, iterative learning, learning curve, mathematical modeling, fatigue, rest, skill acquisition, forgetting.

СУЩНОСТЬ, ПРИЗНАКИ И ПРИНЦИП ОПИСАНИЯ ОРГАНИЗАЦИОННЫХ СИСТЕМ[#]

Г. А. Угольницкий

Южный федеральный университет, г. Ростов-на-Дону

✉ gaugolnickiy@sfedu.ru

Аннотация. Обсуждаются понятие и сущность организационной системы. Общая характеристика этого понятия дана в рамках теории управления организационными системами. В настоящей работе указаны признаки, при наличии которых систему можно отнести к классу организационных (принцип описания организационных систем). Представлена математическая модель организационной системы общего вида, дан иллюстративный пример на основе дуополии Курно. Приведены примеры частных классов организационных систем, специфицированных на основе предложенного принципа. К таким классам отнесены организационно-технические системы специального назначения, организационные системы массового обслуживания, эколого-экономические организационные системы. Для типичных представителей каждого из указанных классов приведено описание в соответствии с предложенной общей моделью. Статья имеет методическую направленность и призвана стандартизовать подходы к описанию организационных систем.

Ключевые слова: организационные системы, теоретико-игровые модели, управление активными агентами.

ВВЕДЕНИЕ

Актуальность проблем организационного управления не вызывает сомнений, поскольку именно качество этого управления определяет производительность труда, экономический рост и тем самым возможности государства и благосостояние населения.

В Философском энциклопедическом словаре приводится следующее определение *организации*: «1) внутренняя упорядоченность, согласованность взаимодействия более или менее дифференцированных и автономных частей целого, обусловленная его строением; 2) совокупность процессов или действий, ведущих к образованию и совершенствованию взаимосвязей между частями целого; 3) объединение людей, совместно реализующих некоторую программу или цель и действующих на основе определенных процедур и правил» [1]. Со-

вокупность этих процедур и правил называется *механизмом функционирования*. Третье значение термина «организация» и есть определение *организационной системы*. В рамках методологии используется аналогичное определение [2]: *организация* – комплексная деятельность с целью создания внутренней упорядоченности, согласованности взаимодействия более или менее дифференцированных и автономных элементов предмета этой деятельности (в том числе путем образования и поддержания между этими элементами взаимосвязей с заданными характеристиками).

Формальные модели управления организациями предоставляют теории активных систем и управления организационными системами (ОС) [3, 4], информационная теория иерархических систем [5–8], теория контрактов и дизайн механизмов [9], теория управления устойчивым развитием активных систем [10–12]. В работе [13] показаны возможности использования искусственных нейронных сетей в исследовании иерархических организационных систем. В значительной степени

[#] Работа выполнена при финансовой поддержке Российского научного фонда, проект №25-11-00094.



следует согласиться с мнением автора [13] о том, что «многошаговые иерархические игры многих лиц представляют собой **ЯЗЫК** теории управления организационными системами», хотя модели теории управления организационными системами используют и другие математические конструкции. Согласно фундаментальной монографии [4] объектом исследований теории управления организационными системами (ТУОС) являются сами ОС, предметом исследований – механизмы управления, а основным методом исследования – математическое моделирование (с. 7). В соответствии с этим подходом основное внимание в работе [4] уделяется задачам и механизмам управления, а именно, выделяются управление составом и структурой ОС, институциональное управление ограничениями и нормами деятельности, мотивационное управление интересами и предпочтениями, информационное управление, управление порядком функционирования – последовательностью получения информации и выбора стратегий агентами (с. 9). Наконец, модель ОС определяется заданием состава ОС (множества её участников), структуры ОС (совокупности различных связей между участниками), множества допустимых стратегий участников ОС, их предпочтений, информированности и порядка функционирования (с. 8).

Вместе с тем в значительной степени остаётся открытым вопрос о водоразделе между подходами менеджмента (который было бы правильно перевести на русский язык именно как управление организациями) и собственно ТУОС. Заслуживает внимания и чёткий ответ на вопрос о том, какую систему можно считать организационной (ведь не все сложные системы реального мира являются таковыми).

Вклад настоящей статьи состоит в:

- описании сущности организационных систем и предложении их математической формализации, определяющей подход ТУОС;
- указании признаков, при наличии которых систему можно отнести к классу организационных (принцип описания организационных систем);
- подробном описании ряда частных классов организационных систем, выделенных в соответствии с указанным принципом.

В § 1 описывается сущность организационных систем, их признаки и общая математическая модель. В § 2 рассматривается принцип описания классов организационных систем. В § 3 приводятся примеры специфических классов организационных систем, полученных применением этого принципа. Итоги и перспективы обсуждаются в заключении.

1. СУЩНОСТЬ И ПРИЗНАКИ ОРГАНИЗАЦИОННЫХ СИСТЕМ

В замечательной во многих отношениях книге [14] дано следующее определение: «Социальная система – это динамическая совокупность автономных индивидов, которые преследуют собственные цели во взаимодействии с окружающей средой» (с. 20). Это определение довольно близко к тому, что следует понимать под ОС. Однако, поскольку любое сложное понятие есть синтез многих определений, то более удобным представляется развёрнуто определить ОС через её математическую модель. Эта модель синтезирует приведённое во введении описание из работ [3, 4] и модель активной сети из статьи [15] с некоторыми существенными уточнениями и дополнениями. Считаем, что ОС состоит из управляющей (активной) подсистемы и управляемой подсистемы (отметим, что такая конструкция включает и организационно-технические системы). Предлагаемая модель ОС имеет вид

$$\langle N, A, X, I, U, S, F, J, R \rangle. \quad (1.1)$$

Здесь:

$N = \{0, 1, \dots, n\}$ – множество активных агентов, образующих состав управляющей подсистемы ОС. Индексом 0 обозначается выделенный агент (Центр), который пока считается единственным.

$A = \{(i, j)\}$ – множество связей различного типа между активными агентами, $i, j \in N$. Ориентированный граф $D = (N, A)$ определяет структуру связей ОС, которые определяются отношениями субординации, потоками передачи вещества, энергии и информации.

X – множество состояний управляемой подсистемы ОС (подмножество некоторого топологического векторного пространства).

$I = I_0 \times I_1 \times \dots \times I_n$, где I_i – доступная i -му агенту информация об ОС и её внешней среде. Поскольку рассматривается детерминированная модель (1.1), то, в отличие от работы [16], эта информация содержит сведения агента о действиях других агентов и их функциях выигрыша, а также о состоянии управляемой подсистемы. Все агенты, включая Центр, стремятся максимизировать свой выигрыш, а при наличии неоднозначности руководствуются принципом гарантированного результата [8].

$U = U_0 \times U_1 \times \dots \times U_n$, где U_i – множество допустимых действий i -го агента (подмножество некоторого конечномерного пространства [8]).

$S = S_0 \times S_1 \times \dots \times S_n$, где S_i – множество допустимых стратегий i -го агента. Допустимая стратегия $s_i \in S_i$ есть отображение $s_i: I_i \rightarrow U_i$, которое определяет выбор i -м агентом допустимого действия в зависимости от имеющейся у него информации.

F – правило изменения состояний с учётом действий активных агентов. Это может быть система алгебраических, дифференциальных или разностных уравнений либо алгоритм, явно определяющий переходы между состояниями управляемой подсистемы. Это правило можно рассматривать как обобщённый оператор, действующий в пространстве состояний.

$J = (J_0, J_1, \dots, J_n)$ – набор функционалов выигрыша активных агентов. Отображение $J_i: U \times X \rightarrow \mathbb{R}$ определяет выигрыш i -го агента в зависимости от действий всех агентов и текущего состояния управляемой подсистемы. Критерием эффективности организации в целом является значение функционала выигрыша Центра, зависящее в общем случае от действий всех участников и состояния управляемой подсистемы.

R – регламент функционирования ОС, алгоритмически определяющий порядок выбора активными агентами стратегий, возможную передачу информации другим агентам и изменение состояния управляемой подсистемы.

Соответственно, можно выделить следующие признаки ОС.

1. *Активность агентов.* Каждый агент имеет свой функционал выигрыша J_i и самостоятельно выбирает допустимую стратегию $s_i \in S_i$. В рамках модели оптимизация функционала выигрыша полностью определяет интересы и предпочтения агента. В частности, агенты могут сознательно искажать передаваемую другим агентам информацию в своих интересах (так называемая проблема *манипулируемости* процедур принятия решений [17]). Другими проявлениями активности могут быть *дальновидность* поведения агентов [18] и наличие у них *рефлексии* относительно своей деятельности и деятельности других агентов [16].

2. *Целеполагание.* Организационная система имеет некоторую цель, устанавливаемую самостоятельно или задаваемую извне (обществом, вышестоящей организацией). В общем случае эта цель, выражаемая в виде ограничения, состоит как минимум в выполнении условия жизнеспособности организации $X \subseteq X^*$, при котором значения всех существенных показателей её функционирования принадлежат заданному диапазону. Технически

можно включить требование жизнеспособности в функционал выигрыша Центра J_0 посредством штрафной функции $\rho(X, X^*)$, которая равна нулю при выполнении условия жизнеспособности и принимает бесконечно большое значение при его нарушении. Если условие жизнеспособности отсутствует, то цель ОС состоит только в максимизации функционала J_0 без штрафной функции.

3. *Организация.* Организационная система (расширенная активная система) образована управляющей подсистемой, состоящей из активных агентов, включая Центр, и управляемой подсистемой. В состав управляемой подсистемы активные агенты не входят: она включает в себя технические, экономические и иные составляющие, контролируемые активными агентами. Взаимодействие между активными агентами устанавливается регламентом R и определяет динамику управляемой подсистемы (изменение её состояния $x \in X$ во времени в силу правила F) и выигрыши агентов J_i .

2. ПРИНЦИП ОПИСАНИЯ ОРГАНИЗАЦИОННЫХ СИСТЕМ. ПРИМЕР

Чтобы назвать некоторую систему организационной, необходимо и достаточно, чтобы она имела все три перечисленных выше признака ОС: активность агентов, целеполагание, механизм управления. Тогда можно построить модель (1.1) этой ОС и указать конкретные значения её составляющих. В частных случаях некоторые составляющие могут быть пустыми множествами или принимать тривиальные значения.

Рассмотрим следующий упрощённый иллюстративный пример (игра Гермейера Γ_2 в дуополии Курно [8]). Здесь в выражении (1.1) $N = \{0, 1\}$. Организация состоит из двух активных элементов: Центра и агента; $A = \{(0, 1)\}$. Центр делает первый ход, формулируя «правила игры» для агента.

Центр знает U_0, U_1, J_0, J_1 , будет иметь информацию о выборе $u_1 \in U_1$ и знает, что выбор агентом стратегии определяется его стремлением к максимизации своего выигрыша, то есть стратегия Центра (механизм управления) – отображение множества U_1 в U_0 . Агент знает U_1 и J_1 . При известной стратегии агента Центр стремится максимизировать свой выигрыш J_0 . При неоднозначности выбора агента Центр руководствуется обобщённым принципом гарантированного результата Ю. Б. Гермейера [5, 6, 8]. Имеем



$$J_0 = (a - c_1 - u_1 - u_2)u_1,$$

$$J_1 = (a - c_2 - u_1 - u_2)u_2,$$

$$c_1 < c_2 < a.$$

Здесь $U_0 = U_1 = [0, 1]$, $s_0 : U_1 \rightarrow U_0$, $s_1 \equiv u_1$. Модель статическая, управляемая подсистема отсутствует, т. е. множество X и правило F не рассматриваются. Цель организации – максимизация выигрыша Центра J_0 .

Центр делает первый ход: выбирает стратегию

$$u_0(u_1) = \begin{cases} (a - c_1) / 2, & u_1 = 0, \\ a - c_1 - u_1, & u_1 > 0, \end{cases}$$

и сообщает её агенту. Единственный оптимальный ответ агента $u_1 = 0$, в результате чего

$$J_0 = (a - c_1)^2 / 4, J_1 = 0 \text{ (в противном случае } J_1 < 0).$$

Возможная содержательная интерпретация для дуополии Курно: фирма 0 предлагает фирме 1 некоторые условия, становясь лидером дуополии. Скажем, фирма 0 может предложить фирме 1 побочный платёж за более высокую цену на тендере с участием фирмы 0. Более точно эта ситуация описывается дуополией Бертрана [7], но объёмная конкуренция по Курно также подходит для её описания.

3. ПРИМЕРЫ

Приведём несколько примеров, иллюстрирующих предложенный подход к описанию ОС. Сначала даётся общая характеристика некоторого класса ОС. Затем она конкретизируется для ОС определённого вида, принадлежащего выделенному классу.

3.1. Организационно-технические системы специального назначения

Рассмотрим в качестве первого развёрнутого примера ОС организационно-технические системы специального назначения (ОТССН). Здесь речь идёт о системах наблюдения и перехвата летательных аппаратов и ракет противника. Характеристика таких систем и их примеры даны в статье [19]. В общем случае составляющие ОТССН можно охарактеризовать следующим образом.

N – группа агентов-наблюдателей. Роль Центра играет командир группы. Заметим, что сторону противника тоже представляют активные агенты. Они не включаются в состав ОС (являясь для нее внешней средой), но их действия могут учитываться в модели ОС, в том числе при определении состояния её управляемой подсистемы и в ряде других аспектов, например, активного противоборства.

A – множество связей между агентами, определяемых необходимостью оперативного взаимодействия в процессе наблюдения и перехвата.

I_i – информация, которой располагает i -й агент относительно действий остальных агентов, Центра и противника, а также состояния управляемой системы.

U_i – множество допустимых действий i -го агента (выбор позиции, параметры наведения зенитных устройств).

S_i – множество стратегий i -го агента, зависящих от действий противника. Поскольку эти действия влияют на состояние системы, то в общем случае используются позиционные стратегии, зависящие не только от времени, но и от состояния системы.

$x = (x_1, \dots, x_m) \in X$ – вектор состояния системы (координаты и скорости наблюдаемых объектов).

F – общее правило изменения состояния системы, определяемое механическими и геометрическими условиями и закономерностями её функционирования.

J_i – функционал выигрыша i -го агента, определяемый точностью обнаружения цели.

R – порядок принятия решений в ОТССН. Он определяется регламентом некоторой игры Гермейера [8], где в роли ведущего выступает Центр, а в роли одновременно выбирающих действия ведомых – агенты-наблюдатели и агенты противника.

Возьмём в качестве конкретного примера ОТССН триангуляционную измерительную систему (ТИС) [19]. В этом случае множество агентов N образуют операторы ТИС, каждый из которых ассоциируется со своим измерительным пунктом. Множество A задаётся парными двусторонними связями между агентами и Центром. По этим связям агенты передают Центру информацию о своих действиях и о состоянии управляемой системы, а Центр сообщает агентам выбранные им стратегии. Каждый агент знает множество своих действий и функционал выигрыша, что определяет множество I_i . Задача управления рассматривается в статической постановке, поэтому вектор состояния x и правило F не учитываются.

Пусть $\{P_n\}_{n=1}^N = \{[x_n^p, y_n^p]\}_{n=1}^N$ – измерительные

пункты ТИС, $\{S_m\}_{m=1}^M = \{[x_m^s, y_m^s]\}_{m=1}^M$ – пункты системы

постановки помех. Здесь N – общее количество измерительных пунктов, M – общее количество пунктов постановки помех, x и y – координаты пунктов на плоскости. Верхний индекс p – измерительный пункт, s – пункт постановки помех. Нижний индекс – номер пункта. Задачу наблюдателя можно записать следующим образом:

$$J(u_p, u_s) = |N - K| \rightarrow \max_{u_p},$$

$$K < \lfloor N / 2 \rfloor + 1,$$

$$\forall i, j \in \{1, \dots, N\} : \|P_i - P_j\| > B_{\min},$$

$$\forall i, j, k \in \{1, \dots, N\} : \frac{y_k^p - y_i^p}{y_j^p - y_i^p} \neq \frac{x_k^p - x_i^p}{x_j^p - x_i^p},$$

где K – количество измерительных пунктов, находящихся в зоне помех; $u_p = [P_1 \dots P_N]$,

$u_s = [\alpha_1, S_1, \dots, \alpha_M, S_M]$; $\lfloor \cdot \rfloor$ – округление до целого в меньшую сторону; $\|\cdot\|$ – евклидова норма; α_m – угол поворота сектора помех m -го пункта. Здесь выполнение условия $K < \lfloor N/2 \rfloor + 1$ обеспечивает необходимое для функционирования ТИС количество рабочих измерительных пунктов (более половины общего количества). Оставшиеся два ограничения – это требования к топологии ТИС: $B_{\min}$ – минимально допустимое расстояние между пунктами ТИС; никакие три измерительных пункта ТИС не должны лежать на одной прямой. Итак, множество допустимых действий U_i определено; для простоты в работе [19] считается, что стратегии S_i совпадают с действиями.

Значение целевой функции $J(u_p, u_s)$ представляет собой количество рабочих измерительных пунктов. Наблюдатель максимизирует это количество с учётом указанных ограничений. Задача противника состоит в её минимизации по u_s :

$$J(u_p, u_s) = |N - K| \rightarrow \min_{u_s},$$

$$K \geq \lfloor N/2 \rfloor + 1.$$

Таким образом, функционалы выигрыша $J = J_1 = -J_2$ определяют антагонистическую игру между наблюдателем и противником. Условие $K \geq \lfloor N/2 \rfloor + 1$ требует, чтобы для значения, обеспечивающего минимум целевой функции, количество измерительных пунктов ТИС, которым поставлена помеха, превышало половину от общего их числа (только в этом случае ТИС становится неработоспособной).

Порядок принятия решений R следующий: каждый агент выбирает свои допустимые действия и сообщает их Центру. Центр анализирует ситуацию в целом и утверждает предложенные агентами действия либо корректирует их в форме приказа, после чего множество Λ может рассматриваться как единый игрок. Противник действует аналогично, тем самым определяются исход антагонистической игры и выигрыши игроков [19].

3.2. Организационные системы массового обслуживания

В качестве второго развёрнутого примера ОС возьмём организационные системы массового обслуживания (ОСМО). Этот класс ОС введён в серии работ [20–24]. В этом случае:

N – множество активных агентов, осуществляющих обслуживание потока заявок. В роли Центра выступает руководство организации, обеспечивающей обслуживание.

A – множество связей между агентами, возникающих в процессе обслуживания.

I_i – информация, которой располагает i -й агент относительно целей и возможностей остальных агентов

(особенно Центра) и принципов принятия ими решений.

U_i – множество допустимых действий i -го агента, связанных с процессом обслуживания (значения интенсивности и качества обслуживания).

Центр назначает дисциплину обслуживания, определяет мощность обслуживающих устройств (например, через величину расходов на их закупку и сопровождение), состав, структуру и квалификацию агентов, значения параметров административных и экономических механизмов управления агентами.

S_i – множество стратегий i -го агента. Как и в случае ОТССН, стратегии Центра – это механизмы управления с обратной связью по действиям агентов. Однако возможны и вырожденные механизмы управления без обратной связи, например, нормативные установления. Стратегии остальных агентов могут как совпадать с их действиями (программные стратегии), так и зависеть от состояния управляемой системы (позиционные стратегии) или действий других агентов (стратегии с обратной связью по управлению).

$x = (x_1, \dots, x_m) \in X$ – вектор состояния ОСМО, ко-

торый включает в качестве основных переменных интервал времени между последовательно поступающими заявками на обслуживание и время обслуживания одной заявки, m – число рассматриваемых переменных состояния. Эти переменные могут дифференцироваться для различных обслуживающих устройств. Производные переменные – время ожидания заявки и время простоя обслуживающего оборудования. Возможно также введение дополнительных переменных, отражающих специфику конкретной ОСМО.

F – общее правило изменения состояния ОСМО. Оно определяется заданием законов распределения упомянутых выше случайных величин и их параметров.

J_i – функционал выигрыша i -го агента, равный разности между его доходом и затратами на трудовые усилия и повышение квалификации. Принципиальная цель Центра в ОСМО заключается в обеспечении баланса между временем ожидания заявки в очереди и временем простоя обслуживающего оборудования, который определяет жизнеспособность ОСМО.

R – порядок принятия решений в ОСМО. Основную роль здесь играют механизмы административного и экономического управления Центра, зная которые, свои решения принимают агенты обслуживания.

Теперь рассмотрим ОСМО вида «железнодорожная станция – морской порт» [20–24]. Схема такой ОСМО показана на рис. 3.1. В качестве органа внешнего управления (Центра) может выступать Министерство транспорта РФ, региональные министерства транспорта или иной орган управления, способный координировать работу железной дороги и морского порта по распоряжению правительства или по добровольному согласию экономических субъектов. В роли активных подсистем (агентов) выступают руководство и персонал железнодорожной станции и морского порта, что определяет множество N . Множество A показано на рис. 3.1.

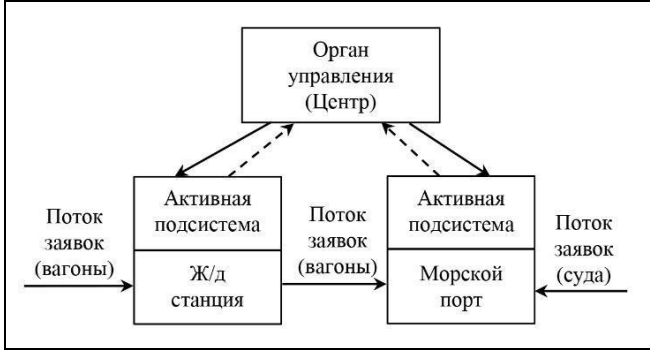


Рис. 3.1. ОСМО «железнодорожная станция – морской порт» [24]

Как видно из рис. 3.1, данная ОСМО содержит две подсистемы массового обслуживания: «вагон – порт» (левая часть рисунка) и «судно – порт» (правая часть рисунка). Рассмотрим сначала подсистему «вагон – порт». Каждый вагон будем характеризовать набором $w = \langle y, g, s, j, m, b, \omega \rangle$,

$$y \in Y, g \in G, s \in S, j \in J, m \geq 0, b \in B, \omega \in \Omega,$$

где y – тип вагона, Y – множество типов вагонов, рассматриваемых в рамках данной модели; g – род груза, G – множество родов грузов; s – тип груза, S – множество разных типов грузов; j – номер партии, J – множество номеров партий груза; m – масса нетто груза в вагоне, если вагон поступает на выгрузку, либо масса нетто требуемого к погрузке груза, если вагон поступает на погрузку; b – признак, указывающий на тип грузовой операции с данным вагоном (1 – погрузка, 0 – выгрузка, т. е. $B = \{0, 1\}$); ω – тип поезда, с которым прибыл вагон, Ω – множество рассматриваемых типов поездов.

В качестве требований второй подсистемы «судно – порт» будем рассматривать суда, приходящие в порт для осуществления грузовых операций. Обслуживающие устройства для них – это причалы, оборудованные погрузочно-разгрузочной техникой.

Введём набор характеристик для каждого судна:

$$\eta = \langle \tilde{G}, \tilde{S}, \tilde{J}, \Phi, l \rangle.$$

Здесь $\tilde{G} \subset G$ – множество родов перевозимого груза; $\tilde{S} \subset S$ – множество типов перевозимого груза; $\tilde{J} \subset J$ – множество номеров партий груза; l – длина судна; $\Phi: \tilde{G} \times \tilde{S} \times \tilde{J} \rightarrow \mathbb{R}^+ \times B \times \mathbb{N}$ – отображение, задающее структуру загрузки судна, т. е. сопоставляющее каждой тройке (g, s, j) тройку (m, b, d) , где $m \in \mathbb{R}^+$ – масса груза g -го рода s -го типа j -й партии на судне к выгрузке (погрузке), $b \in B$ – тип грузовой операции (погрузка или выгрузка), $d \in \mathbb{N}$ – натуральное число, задающее порядок осуществления грузовой операции с данным грузом. Подробнее переменные состояния x и правило их изменения F описаны в работах [20–23].

Динамическая модель иерархического управления ОСМО имеет вид [21–23]:

$$H(s_1, s_2) = \int_0^T [k_1 s_1(t) V(y_1(t)) + k_2 s_2(t) y_2(t) - ds_1^\gamma(t) - ds_2^\gamma(t)] dt, \quad (3.1)$$

$$C_1(y_1, s_1) = \int_0^T \left[c_1 y_1^{\frac{1}{\alpha}}(t) - k_1 s_1(t) V(y_1(t)) - c_3 x_1(t) p_1 - P \right] dt, \quad (3.2)$$

$$C_2(y_1, s_1) = \int_0^T \left[c_2 y_2^{\frac{1}{\beta}}(t) - k_2 s_2(t) y_2(t) - c_4 x_1(t) p_2 - k_3 Q(y_2(t)) \right] dt, \quad (3.3)$$

$$s_1 = (s_1(t_0), s_1(t_1), \dots, s_1(T)), \quad (3.4)$$

$$s_2 = (s_2(t_0), s_2(t_1), \dots, s_2(T)), \quad (3.5)$$

$$c_1 = (c_1(t_0), c_1(t_1), \dots, c_1(T)), \quad (3.6)$$

$$s_1^{\min}(t) \leq s_1(t) \leq s_1^{\max}(t), \quad (3.7)$$

$$y_1^{\min}(t) \leq y_1(t) \leq y_1^{\max}(t), \quad (3.8)$$

$$s_2^{\min}(t) \leq s_2(t) \leq s_2^{\max}(t), \quad (3.9)$$

$$y_2^{\min}(t) \leq y_2(t) \leq y_2^{\max}(t), \quad (3.10)$$

$$x_1(t) = x_1(t-1) + y_1^{\max}(t) - y_2(t), \quad (3.11)$$

$$x_2(t) = x_2(t-1) + y_1(t) - y_2(t-1); \quad (3.12)$$

$$V(y_1(t)) = \begin{cases} \frac{y_1(t)}{w_1}, & \text{mod}(y_1, w_1) = 0, \\ \frac{y_1(t)}{w_1} + 1, & \text{mod}(y_1, w_1) > 0, \end{cases} \quad (3.10)$$

$$Q(y_2(t)) = \begin{cases} \frac{y_2(t)}{w_2}, & \text{mod}(y_2, w_2) = 0, \\ \frac{y_2(t)}{w_2} + 1, & \text{mod}(y_2, w_2) > 0. \end{cases} \quad (3.11)$$

Здесь H, C_1, C_2 – функции прибыли Центра и агентов за время t ; $s_1(t)$ – управление Центра агентом 1 в момент времени t , т. е. цена аренды стыковочного места для клиентов РЖД; $s_2(t)$ – управление Центра агентом 2, т. е. размер портового сбора, необходимый для уплаты при перевозке товаров; $y_1(t)$ – управление агента 1 (железнодорожная транспортная компания), т. е. размер продукции, отправляемой со склада компании в порт; $y_2(t)$ – управление агента 2 (порта), т. е. объём продукции, отправляемой в качестве поставки со склада порта; $x_i(t)$ – количество груза, не доставленного конечному потребителю агентом i ; $c_i(t)$ – штрафной коэффициент; p_1, p_2 – ценовые коэффициенты; P – дополнительные затраты; α, β, γ – подгоночные параметры;

k_1, k_2, k_3 – коэффициенты для согласования размерности [20–22].

Модель (3.1)–(3.11) содержит описание элементов J_i и U_i общей модели (1.1). Порядок функционирования R следующий: Центр выбирает свои управляющие воздействия $s_1(t), s_2(t)$ и сообщает их агентам. Зная управления Центра, агенты одновременно и независимо выбирают значения своих управляющих воздействий $y_1(t), y_2(t)$ соответственно. Таким образом, стратегии агентов и Центра совпадают с их действиями (динамическая игра Гермейера Γ_{II}), взаимная информированность игроков I_i также описана.

3.3. Эколого-экономические организационные системы

Третий пример специфического класса ОС образуют эколого-экономические организационные системы (ЭЭОС). Механизмы управления ЭЭОС, включающие учёт специфики той или иной управляемой подсистемы, описаны в монографиях [25, 26] и статьях [27–30]. В общем виде составляющие ЭЭОС можно охарактеризовать следующим образом.

N – множество агентов (природопользователей), ассоциированных с экосистемой, образующей природную основу рассматриваемой эколого-экономической системы. Сюда относятся находящиеся на занимаемой экосистемой территории (прилегающие к акватории данной экосистемы) предприятия промышленности, сельского хозяйства, рекреации, транспорта и иных отраслей хозяйства, а также население. Роль Центра играет администрация территории.

A – множество транспортных, хозяйственных и административных связей между агентами.

I_i – информация, которой располагает i -й агент относительно действий и функций выигрыша остальных агентов (особенно Центра).

U_i – множество допустимых действий i -го агента. Центр назначает допустимые квоты эксплуатации природных ресурсов (вылов рыбы, рубка леса, добыча полезных ископаемых и т. п.) и предельно допустимые концентрации (выбросы) загрязняющих веществ в окружающей среде, а также значения параметров механизмов экономического регулирования (налоги, штрафы, льготы, субсидии). Хозяйствующие агенты выбирают объёмы производства, цены на продукцию, параметры природоохранной деятельности.

S_i – множество стратегий i -го агента. Как правило, стратегии Центра – это механизмы управления с обратной связью по действиям агентов. В частности, при учёте оппортунистического поведения в системе управления возникает обратная связь по величине взятки Центру от агентов, позволяющая, например, ослаблять экологические требования или выделять дополнительные ресурсы. Однако возможны и механизмы управления без обратной связи, например, законодательные установления. Стратегии остальных

агентов могут как совпадать с их действиями (программные стратегии), так и зависеть от состояния управляемой системы (позиционные стратегии) или действий других агентов (стратегии с обратной связью по управлению).

$x = (x_1, \dots, x_m) \in X$ – вектор состояния эколого-экономической системы. Его компоненты естественно разбить на два подмножества, описывающие соответственно экологическую и экономическую подсистемы ЭЭОС.

F – общее правило изменения состояния ЭЭОС. Фактически оно тоже состоит из двух частей: во-первых, материальные балансы и производственные функции, описывающие экономическую деятельность, во-вторых, правила изменения значений природных показателей. В свою очередь, вторая часть подразделяется на описание естественной динамики природной среды и её антропогенного изменения (загрязнение, эксплуатация природных ресурсов, охрана окружающей среды и т. д.).

J_i – функционал выигрыша i -го агента. Здесь в качестве активных агентов рассматриваются хозяйствующие субъекты, для которых выигрыш – это прибыль с учётом природоохранных затрат и возможных экологических штрафов. Центр имеет два критерия: экономический рост на подведомственной территории и соблюдение экологических требований.

R – порядок принятия решений в ЭЭОС. Основную роль здесь играют механизмы административного и экономического управления Центра, который сообщает их агентам. Зная эти механизмы, свои решения принимают хозяйствующие субъекты.

Теперь рассмотрим ЭЭОС, возникающую при управлении рыболовством [31]. Здесь активные агенты – рыбохозяйственные предприятия, Центр – природоохранный орган. Связи ограничиваются воздействием Центра на агентов. Агенты максимизируют целевые функционалы

$$J_i = \int_0^T e^{-\rho t} \{a v_i(t) P(t) - s_i(t) M[P(t) - P^*]^2\} dt - e^{-\rho T} s_i(T) M[P(T) - P^*] \rightarrow \max \quad (3.12)$$

при ограничениях на управления (допустимые действия агентов)

$$q_i(t) \leq u_i(t) \leq r_i, \quad i = 1, \dots, N, \quad (3.13)$$

и уравнениях динамики биомассы рыбной популяции (правило изменения F скалярной переменной состояния $x = P$)

$$\begin{aligned} \frac{dP}{dt} = & [\varepsilon - \beta P(t) + \alpha \left(\sum_i u_i(t) \right)^\gamma - \\ & - \sum_i \alpha_i (r_i - u_i(t))^{\gamma_i}] P(t), \quad (3.14) \\ P(0) = & P_0, \quad i = 1, \dots, N. \end{aligned}$$



Здесь P_0 , $P(t)$ – начальное и текущее значения биомассы рыбной популяции; P^* – его идеальное значение, полностью удовлетворяющее требованиям жизнеспособности популяции. При нарушении условия жизнеспособности на агентов налагается штраф с коэффициентом $M \gg 1$; a – цена единицы биомассы рыбы; ε – коэффициент естественного прироста рыбной популяции, β – коэффициент самолимитирования; ρ – коэффициент дисконтирования; α, γ – подгоночные параметры модели.

В работе [31] рассматривается также упрощённая линейная версия модели (3.12)–(3.14).

Предполагается, что каждый агент распределяет свой ресурс между общественными и частными интересами, поэтому его выигрыш складывается из двух составляющих – дохода от частной деятельности и доли ущерба от общественного зла, с которым совместно борются агенты. Агенты (рыболовецкие предприятия) $i = 1, \dots, N$ максимизируют доход от рыболовства с учётом возможного штрафа за нарушение условий жизнеспособности рыбной популяции. Условие жизнеспособности (цель управления Центра) имеет вид $\forall t P(t) = P^*$ или, в более слабой форме, $\forall t [P(t) - P^*]^2 \leq \delta$.

Управление агента $u_i(t)$ – часть ресурса r_i , ассигнуемая на общественные нужды (тогда $r_i - u_i(t)$ – часть ресурса, выделяемая на частную деятельность). В данной модели $r_i - u_i(t)$ – это инвестиции в наращивание промысловых усилий, тогда доля вылова рыбы i -м предприятием вычисляется как функция промыслового усилия $v_i(t) = h_i(r_i - u_i(t))$. Без существенного ограничения общности $v_i(t) = k_i(r_i - u_i(t))^{p_i}$, $0 < p_i < 1$, k_i, p_i – подгоночные параметры модели. Величина $u_i(t)$ – ассигнования в повышение экологичности промысла и рыбозаведение.

Модель (3.12)–(3.14) представляет собой дифференциальную игру N лиц, в которой условия жизнеспособности учтены с помощью штрафов $M[P(t) - P^*]^2$ в целевых функционалах. Переменные $s(t) = \{s_i(t)\}_{i=1}^N \in S$ естественно трактовать как экономические управления (управления побуждения) Центра как государственного контрольного органа верхнего уровня (например, службы рыбного хозяйства) такие, что

$$0 \leq s_i(t) \leq 1, \sum_{i=1}^N s_i = 1; t \geq 0, i = 1, \dots, N. \quad (3.15)$$

Центр может использовать и административное управление (принуждение) $q(t) = \{q_i(t)\}_{i=1}^N \in Q = \{0 \leq q_i(t) \leq r_i, t \geq 0\}$, выбирая значения переменных $q_i(t)$ из условия

$$0 \leq q_i(t) \leq r_i, t \geq 0, i = 1, \dots, N. \quad (3.16)$$

Считается, что интересы Центра описываются стремлением к максимизации функционала (утилитаристской функции общественного благосостояния)

$$J = \sum_{i=1}^N J_i \rightarrow \max. \quad (3.17)$$

Тогда модель (3.12)–(3.17) представляет собой иерархическую дифференциальную игру Центра с несколькими активными агентами нижнего уровня управления. Примем следующие предположения относительно информационного регламента (порядка R) такой игры: все игроки используют программные стратегии; Центр выбирает и сообщает агентам экономические (3.15) либо административные управления (3.16), которые могут быть только функциями времени либо зависеть также от управлений агентов; в частности, Центр назначает штрафы; при известных стратегиях Центра агенты одновременно и независимо выбирают свои действия, что приводит к равновесию Нэша в игре агентов в нормальной форме [31]. Все игроки, включая Центр, знают свои множества стратегий и функционалы выигрыша; знание игроков о множествах стратегий и функционалах выигрыша других игроков определяется регламентом некоторой игры Гермейера [6–8].

ЗАКЛЮЧЕНИЕ

Традиционный менеджмент как наука и теория управления организационными системами имеют один и тот же объект – управление организациями. Однако предметом менеджмента являются организационно-экономические и социально-психологические отношения и методы управления, а предмет ТУОС – математические модели, информационные технологии и методы их применения в управлении организациями. Подчеркнём, что математические модели в ТУОС используются исключительно как инструменты для решения задач (построения и совершенствования механизмов) управления, а не как подлежащие исследованию абстрактные формально-логические конструкции чистой математики. Отметим, кстати, что с точки зрения классификатора ВАК первому направлению соответствует специальность 5.2.6 Менеджмент (экономические науки), а второму – специальность 2.3.4 Управление в организационных системах (технические науки).

В настоящей статье предложена общая математическая модель, описывающая организационную систему с точки зрения ТУОС. Построение такой модели и её применение для анализа и управления конкретной организацией или классом организаций позволяет отнести исследование к ТУОС (принцип описания ОС). Приведены развёрнутые примеры частных классов ОС и их конкретных представителей (ОС специального назначения, ОС массового обслуживания, эколого-экономические ОС), описанных в терминах предложенной математической модели. Конечно, круг известных на

сегодняшний день моделей управления ОС, включающих помимо активных агентов и управляемую ими подсистему той или иной природы, гораздо шире. Здесь следует упомянуть производственные системы [32,33], организационно-технические системы [33], управление проектами [33], вооружённую борьбу [34] и др.

Для простоты изложения модель приведена в детерминированной постановке. Конечно, на самом деле в функционировании ОС существенную роль играют неопределённые факторы различной природы. С точки зрения рассматриваемой в статье проблематики учёт неопределённости будет лишь усложнять понимание. Однако учёт неопределённости, несомненно, формирует первое важное направление дальнейшего развития исследований.

Вторым естественным направлением уточнения предложенной модели является учёт ограниченной рациональности активных агентов (Г. Саймон, Р. Хайнер, Р. Зельтен, Д. Канеман, А. Тверски и др.).

Третье направление развития – исследование ОС с несколькими центрами управления (случай распределённого контроля), а также многоуровневых ОС.

Заметим, что управляемая подсистема ОС может иметь техническую, биологическую, экономическую или иную природу. Учёт специфики управляемой подсистемы, определяемой характеристикой её состояния, правилом его изменения и другими элементами модели, играет существенную роль при разработке и реализации механизмов управления ОС. Хотя закономерности процессов управления и структура механизмов управления имеют общий характер, на практике нельзя успешно управлять ОС, не понимая её технико-экономической специфики. Поэтому модельное описание динамики управляемой подсистемы и условий её жизнеспособности играет важную роль в ТУОС наряду с учётом и согласованием интересов входящих в её состав активных агентов.

ЛИТЕРАТУРА

1. *Философский энциклопедический словарь*. – М.: Советская энциклопедия, 1983. – 840 с. [*Filosofskii entsiklopedicheskii slovar'*. – Moscow: Sovetskaya entsiklopediya, 1983. – 840 s. (In Russian)]
2. Белов М.В., Новиков Д.А. Методология комплексной деятельности. – М.: Ленанд, 2018 – 320 с. [*Belov, M.V., Novikov, D.A. Metodologiya kompleksnoi deyatel'nosti*. – M.: Lenand, 2018 – 320 s. (In Russian)]
3. Бурков В.Н., Новиков Д.А. Теория активных систем: состояние и перспективы. – М.: СИНТЕГ, 1999. – 128 с. [*Burkov, V.N., Novikov, D.A. Teoriya aktivnykh sistem: sostoyanie i perspektivy*. – M.: SINTEG, 1999. – 128 s. (In Russian)]
4. Новиков Д.А. Теория управления организационными системами. – М.: Изд-во физ.-мат. лит., 2007. – 584 с. [*Novikov, D.A. Teoriya upravleniya organizatsionnymi sistemami*. – M.: Izd-vo fiz.-mat. lit., 2007. – 584 s. (In Russian)]
5. Гермейер Ю.Б. Введение в теорию исследования операций. – М.: Наука, 1971. – 384 с. [*Germeier, Yu.B. Vvedenie v teoriyu issledovaniya operatsii*. – M.: Nauka, 1971. – 384 s. (In Russian)]
6. Гермейер Ю.Б. Игры с непротивоположными интересами. – М.: Наука, 1976. – 328 с. [*Germeier, Yu.B. Igrы s neprotivopolozhnymi interesami*. – M.: Nauka, 1976. – 328 s. (In Russian)]
7. Кукушкин Н.С., Морозов В.В. Теория неантагонистических игр. – М.: МГУ, 1984. – 106 с. [*Kukushkin, N.S., Morozov, V.V. Teoriya neantagonisticheskikh igr*. – M.: MGU, 1984. – 106 s. (In Russian)]
8. Горелик В.А., Горелов М.А., Кононенко А.Ф. Анализ конфликтных ситуаций в системах управления. – М.: Радио и связь, 1991. – 288 с. [*Gorelik, V.A., Gorelov, M.A., Kononenko, A.F. Analiz konfliktnykh situatsii v sistemakh upravleniya*. – M.: Radioisvyaz', 1991. – 288 s. (In Russian)]
9. Laffont, J.-J., Martimort, D. The Theory of Incentives: The Principal-Agent Model. – Princeton: Princeton University Press, 2002. – 421 p.
10. Угольницкий Г.А. Управление устойчивым развитием активных систем. – Ростов-на-Дону: Изд-во Южного федерального ун-та, 2016. – 940 с. [*Ougolnitsky, G.A. Upravlenie ustoichivym razvitiem aktivnykh sistem*. – Rostov-na-Donu: Izd-vo Yuzhnogo federal'nogo un-ta, 2016. – 940 s. (In Russian)]
11. Угольницкий Г.А. Методология и прикладные задачи управления устойчивым развитием активных систем // Проблемы управления. – 2019. – № 2. – С.19–29. [*Ougolnitsky, G.A. Methodology and Applied Problems of the Sustainable Management in Active Systems*. – Control Sciences. – 2019. – No. 2. – P. 19–29. (In Russian)]
12. Угольницкий Г.А., Горбанёва О.И., Усов А.Б. и др. Теория управления устойчивым развитием активных систем // Управление большими системами. – 2020. – Вып.84. – С.89–113. [*Ougolnitsky, G.A., Gorbaneva, O.I., Usov, A.B., et al. Theory of Sustainable Management in Active Systems. – Large Systems Control*. – 2020. – Iss. 84. – P. 89–113. (In Russian)]
13. Ерешко Ф.И. Приложение иерархических игр к управлению организационными системами // XIV Всероссийское совещание по проблемам управления (ВСПУ-2024): сборник научных трудов / Под общ. ред. Д.А. Новикова. – М.: ИПУ РАН, 2024. – С. 3314–3318. [*Ereshko, F.I. Prilozhenie ierarkhicheskikh igr k upravleniyu organizatsionnymi sistemami* // XIV Vserossiiskoe soveshchanie po problemam upravleniya (VSPU-2024): sbornik nauchnykh trudov / Pod obshch. red. D.A. Novikova. – M.: IPURAN, 2024. – S. 3314–3318. (In Russian)]
14. Словохотов Ю.Л. Физика общества: Применение физических моделей в описании общественных явлений. – М.: ЛЕНАНД, 2024. – 880 с. [*Slovokhotov, Yu.L. Fizika obshchestva: Primenenie fizicheskikh modelei v opisani obshchestvennykh yavlenii*. – M.: LENAND, 2024. – 880 s. (In Russian)]
15. Ougolnitsky, G., Gorbaneva, O. Sustainable Management in Active Networks // Mathematics. – 2022. – Vol. 10, no. 16. – Art. no. 2848.



- 16.Новиков Д.А., Чхартишвили А.Г. Рефлексия и управление: математические модели. – М.: ЛЕНАНД, 2022. – 416 с. [Novikov, D.A., Chkhartishvili, A.G. Refleksiya i upravlenie: matematicheskie modeli. – М.: LENAND, 2022. – 416 s. (In Russian)]
- 17.Бурков В.Н., Еналеев А.К., Коргин Н.А. Согласованность и неманипулируемость механизмов организационного управления: текущее состояние проблемы, ретроспектива, перспективы развития теоретических исследований // Автоматика и телемеханика. – 2021. – № 7. – С. 5–37. [Burkov, V.N., Enaleev, A.K., Korgin, N.A. Incentive Compatibility and Strategy-Proofness of Mechanisms of Organizational Behavior Control: Retrospective, State of the Art, and Prospects of Theoretical Research. – Automation and Remote Control. – 2021. – Vol. 82, no. 7. – P. 1119–1143.]
- 18.Щепкин А.В. Динамические активные системы с дальновидными элементами I. Динамическая модель активной системы // Автоматика и телемеханика. – 1986. – № 10. – С. 89–94. [Shchepkin, A.V. Dinamicheskie aktivnye sistemy s dal'novidnymi elementami I. Dinamicheskaya model' aktivnoi sistemy // Avtomatika i telemekhanika. – 1986. – № 10. – S. 89–94. (In Russian)]
- 19.Угольницкий Г.А., Чепель Е.Н. Теоретико-игровой подход к управлению составом и структурой триангуляционной измерительной системы в условиях априорной неопределённости // Известия РАН. Теория и системы управления. – 2024. – № 2. – С. 53–66. [Ugolnitskiy, G.A., Chepel, E.N. Game-Theoretic Approach to Managing the Composition and Structure of a Bearing-Only Measurement System in Conditions of A Priori Uncertainty. – Journal of Computer and System Sciences International. – 2024. – Vol. 63, no. 2. – P. 343–356.]
- 20.Агиев Х.Р. Модель массового обслуживания в системе логистического взаимодействия железной дороги с морским портом // Инженерный вестник Дона. – 2022. – № 6. – URL: <http://ivdon.ru/magazine/archive/n6y2022/7733>. [Agiev, Kh.R. Model of Queuing in the System of Logistic Interaction of the Railway with the Seaport // Engineering Journal of Don. – 2022. – No. 6. – URL: <http://ivdon.ru/magazine/archive/n6y2022/7733>. (In Russian)]
- 21.Деменский В.И., Мальсагов М.Х., Агиев Х.Р. Модель взаимодействия двух типов транспорта в смешанных перевозках // Инженерный вестник Дона. – 2023. – № 7. – URL: <http://ivdon.ru/magazine/archive/n7y2023/8528>. [Demenskiy, V.I., Mal'sagov, M.K., Agiev, K.R. A Model of Interaction between Two Types of Transport in Multimodal Transport // Engineering Journal of Don. – 2023. – No. 7. – URL: <http://ivdon.ru/magazine/archive/n7y2023/8528>. (In Russian)]
- 22.Агиев Х.Р., Мальсагов М.Х., Угольницкий Г.А. Динамические модели согласования интересов активных участников логистического взаимодействия. Ч. 1 // Системы управления и информационные технологии. – 2024. – № 1 (95). – С. 31–36. [Agiev, Kh.R., Malsagov, M.Kh., Ougolnitsky, G.A., Dynamic Models of the Coordination of Interests of the Active Participants of Logistic Interactions. Part 1 // Sistemy Upravleniya i Informatsionnye Tekhnologii. – 2024. – No. 1 (95). – P. 31–36. (In Russian)]
- 23.Агиев Х.Р., Мальсагов М.Х., Угольницкий Г.А. Динамические модели согласования интересов активных участников логистического взаимодействия. Ч. 2 // Системы управления и информационные технологии. – 2024. – № 3 (97). – С. 30–34. [Agiev, Kh.R., Malsagov, M.Kh., Ougolnitsky, G.A. Dynamic Models of The Coordination of Interests of the Active Participants of Logistic Interactions. Part 2 // Sistemy Upravleniya i Informatsionnye Tekhnologii. – 2024. – No. 3 (97). – P. 30–34. (In Russian)]
- 24.Агиев Х.Р. Организационная система массового обслуживания «железная дорога – морской порт» // Инженерный вестник Дона. – 2025. – № 4. – URL: <http://www.ivdon.ru/magazine/archive/n4y2025/10019>. [Agiev, Kh.R. Queuing Organizational System “Railway - Marine Port” // Engineering Journal of Don. – 2025. – No. 4. – URL: <http://www.ivdon.ru/magazine/archive/n4y2025/10019>. (In Russian)]
- 25.Бурков В.Н., Новиков Д.А., Щепкин А.В. Механизмы управления эколого-экономическими системами / Под ред. акад. С.Н. Васильева. – М.: Изд-во физ.-мат. лит., 2008. – 244 с. [Burkov, V.N., Novikov, D.A., Shchepkin, A.V. Mekhanizmy upravleniya ekologo-ekonomicheskimi sistemami / Pod red. akad. S.N. Vasil'eva. – М.: Izd-vo fiz.-mat. lit., 2008. – 244 s. (In Russian)]
- 26.Угольницкий Г.А. Управление эколого-экономическими системами. – М.: Вуз. книга, 1999. – 132 с. [Ugol'nitskiy, G.A. Upravlenie ekologo-ekonomicheskimi sistemami. – М.: Vuz. kniga, 1999. – 132 s. (In Russian)]
- 27.Мазалов В.В., Реттиева А.Н. Асимметрия в кооперативной задаче управления биоресурсами // Управление большими системами. – 2015. – Вып. 55. – С. 280–325. [Mazalov, V.V., Rettieva, A.N. Asymmetry in a Cooperative Bioresource Management Problem // Large-Scale Systems Control. – 2015. – No. 55. – P. 280–325. (In Russian)]
- 28.Rettieva, A. Cooperation Maintenance in Dynamic Discrete-Time Multicriteria Games with Application to Bioresource Management Problem // J. of Computational and Applied Mathematics. – 2024. – Vol. 441, no. 8. – Art. no. 115699.
- 29.Grune, L., Kato, M., Semmler, W. Solving Ecological Management Problems Using Dynamic Programming // J. of Economic Behavior and Organization. – 2005. – Vol. 57, no. 4. – P. 448–473.
- 30.Fanokoa, P.S., Telahigue, I., Zaccour, G. Buying Cooperation in an Asymmetric Environmental Differential Game // J. of Economic Dynamics and Control. – 2011. – Vol. 35, no. 6. – P. 935–946.
- 31.Сухинов А.И., Угольницкий Г.А., Усов А.Б. Методы решения теоретико-игровых моделей согласования интересов при управлении рыболовством // Математическое моделирование. – 2019. – № 31 (6). – С.127–144. [Sukhinov, A.I., Ougolnitsky, G.A., Usov, A.B. Methods of Solving the Theoretic Game Models for Coordinating Interests in Regulating the Fishery Industry. – Mathematical Models and Computer Simulations. – 2020. – Vol. 12. – P. 176–184.]
- 32.Механизмы управления / Подред. Д.А. Новикова. – М.: Ленанд, 2011. – 192 с. [Mekhanizmy upravleniya / Pod red. D.A. Novikova. – М.: Lenand, 2011. – 192 s. (In Russian)]
- 33.Belov, M.V., Novikov, D.A. Optimal Enterprise: Structures, Processes and Mathematics of Knowledge, Technology and Human Capital. – BocaRaton: CRC Press, 2021. – 344 p.
- 34.Модели военных, боевых и специальных действий / Под ред. Д.А. Новикова. – М.: Ленанд, 2025. – 528 с. [Modeli voennykh, boevykh i spetsial'nykh deistvii / Pod red. D.A. Novikova. – М.: Lenand, 2025. – 528 s. (In Russian)]

Статья представлена к публикации членом редколлегии академиком РАН Д. А. Новиковым.

Поступила в редакцию 27.05.2025,
после доработки 19.06.2025.
Принята к публикации 25.06.2025.

© 2025 г. Угольников Г. А.



Угольников Геннадий Анатольевич – д-р физ.-мат. наук,
Южный федеральный университет, г. Ростов-на-Дону,
✉ gaugolnickiy@sfedu.ru
ORCID iD: <https://orcid.org/0000-0001-5085-5144>

Эта статья доступна по лицензии Creative Commons
«Attribution» («Атрибуция») 4.0 Всемирная.

THE ESSENCE, ATTRIBUTES, AND DESCRIPTION PRINCIPLE OF ORGANIZATIONAL SYSTEMS

G. A. Ougolnitsky

Southern Federal University, Rostov-on-Don, Russia

✉ gaugolnickiy@sfedu.ru

Abstract. The concept and essence of an organizational system are discussed. A general characterization of this concept is provided within the theory of control in organizational systems. Several attributes are indicated to classify a given system as an organizational one (the description principle of organizational systems). A mathematical model of a general organizational system is described, with an illustrative example based on the Cournot duopoly. Extended examples of some classes of organizational systems specified using this principle are presented. Such classes include special-purpose organizational and technical systems, queuing organizational systems, and ecological-economic organizational systems. Typical representatives of each class are described within the general model proposed. This paper has a methodological focus and is intended to define a standard description of organizational systems.

Keywords: organizational systems, game-theoretic models, control in active systems.

Acknowledgments. This work was supported by the Russian Science Foundation, project no. 25-11-00094.

КОМПОЗИЦИИ ДВУХ РАВНОВЕСНЫХ КОДОВ С ОРТОГОНАЛЬНЫМИ ПО ВСЕМ РАЗРЯДАМ КОМБИНАЦИЯМИ ДЛЯ СИНТЕЗА САМОПРОВЕРЯЕМЫХ ДИСКРЕТНЫХ УСТРОЙСТВ

Д. В. Ефанов

Санкт-Петербургский политехнический университет Петра Великого, г. Санкт-Петербург,
Российский университет транспорта (МИИТ), г. Москва

✉ TrES-4b@yandex.ru

Аннотация. Предложено при синтезе контролепригодных и самопроверяемых дискретных устройств использовать композиции двух равновесных кодов с ортогональными по всем разрядам комбинациями. Использование таких кодов позволяет осуществлять контроль вычислений на выходах дискретных устройств по признаку принадлежности не только к кодовым словам равновесных кодов, но и по признаку принадлежности каждой функции, описывающей символы кодовых слов, к классу самодвойственных булевых функций. Показано, за счет чего возможно построение помехозащищенных кодов на основе композиции двух равновесных кодов с ортогональными по всем разрядам комбинациями. Получены формулы, позволяющие определить число ошибок, не обнаруживаемых композициями равновесных кодов по видам (по количеству искажаемых нулей и единиц в кодовых словах) и кратностям. Дана краткая характеристика свойств рассматриваемых кодов. Представлена структура организации схем встроенного контроля для дискретных устройств на основе композиции двух равновесных кодов с ортогональными по всем разрядам комбинациями с контролем вычислений по двум диагностическим признакам. Использование композиции двух равновесных кодов с ортогональными по всем разрядам комбинациями может оказаться эффективным при построении высоконадежных дискретных устройств на различных видах элементной базы.

Ключевые слова: контролепригодные и самопроверяемые устройства, контроль вычислений по двум диагностическим признакам, композиции равновесных кодов, обнаружение ошибок на выходах дискретных устройств.

ВВЕДЕНИЕ

Равновесные коды, или коды « r из n » (r/n -коды), строятся путем выделения из множества кодовых слов с числом символов, равным n (с длиной кодовых слов, равной n), тех кодовых слов, которые имеют один и тот же вес, равный r . Число кодовых слов r/n -кода определяется биномиальным коэффициентом C_n^r . Данные коды описаны в работе [1] и нашли широкое применение в практике обработки и передачи данных [2, 3], а также при синтезе контролепригодных и самопроверяемых

дискретных устройств [4–7]. Теория синтеза детекторов и тестеров равновесных кодов достаточно глубоко проработана [8–12], а особые свойства и характеристики этих кодов исследованы в большом количестве работ, например, в статье [13]. Отметим также, что равновесные коды тесно связаны с теорией комбинаторного блок-дизайна и системами Штайнера [14, 15].

Остановим внимание читателя на применении равновесных кодов при построении устройств с обнаружением неисправностей [16, 17]. При синтезе подобных устройств с использованием равновесных кодов отталкиваются от следующих их

преимуществ. Прежде всего, поскольку кодовые слова данных кодов имеют один и тот же вес, они обнаруживают любые ошибки в кодовых словах, кроме разнонаправленных ошибок с четной кратностью, содержащих группу искажений $\{0 \rightarrow 1, 1 \rightarrow 0\}$ – это так называемые симметричные ошибки. Любые несимметричные ошибки, которые классифицируются на монотонные (связанные с искажениями исключительно нулевых или исключительно единичных значений) и асимметричные (содержащие неравное количество искажений нулей и единиц), они обнаруживают. Эту особенность при синтезе самопроверяемых устройств используют, осуществляя поиск групп монотонно независимых выходов устройства [18, 19] либо монотонно и асимметрично независимых групп выходов [20, 21] и организуя контроль вычислений на них с помощью равновесных кодов. Альтернативным вариантом является модификация структуры контролируемого устройства в структуру с единой группой монотонно независимых выходов либо монотонно и асимметрично независимых выходов [22]. Еще одним свойством равновесных кодов является то, что ими не обнаруживается малое количество ошибок с четной кратностью, что также способствует покрытию большого количества реальных искажений на выходах контролируемых устройств. Третьим важным преимуществом равновесных кодов перед другими является простота контрольного оборудования для них (тестеров, детекторов и блоков контрольной логики) и понятные принципы обеспечения свойства самопроверяемости [9].

При построении самопроверяемых дискретных устройств с использованием равновесных кодов можно применять различные подходы. Интересным является контроль вычислений с использованием не только равновесных кодов, но и свойств булевых функций, описывающих разряды кодовых слов данных кодов, а именно свойств самодвойственных булевых функций. При таком способе организации контроля вычислений существенно увеличиваются показатели контролепригодности в части наблюдаемости [23]. При построении дискретных устройств с контролем вычислений по двум диагностическим признакам можно применить только коды вида $\frac{n}{2}/n$ [24]. В настоящей статье предложено при организации самопроверяемых дискретных устройств использовать композиции пар равновесных кодов с ортогональными

по всем разрядам комбинациями, что существенно расширяет теорию синтеза самопроверяемых дискретных устройств.

1. КОМПОЗИЦИИ ДВУХ РАВНОВЕСНЫХ КОДОВ С ОРТОГОНАЛЬНЫМИ ПО ВСЕМ РАЗРЯДАМ КОМБИНАЦИЯМИ

Среди множества равновесных кодов могут быть выделены такие коды, которые обладают следующим интересным свойством – множество их кодовых слов содержит исключительно пары ортогональных по всем разрядам кодовых слов. Такие коды существуют только для четных значений n , причем для каждого четного n существует

ровно один такой код со значением $r = \frac{n}{2}$. Всё

многообразие равновесных кодов может для наглядности быть показано на треугольнике Паскаля, где каждое число характеризует мощность множества кодовых слов r/n -кодов. Фрагмент треугольника Паскаля изображен на рис. 1. На рис. 1, а показан треугольник с выделенными числами, соответствующими мощностям множеств кодовых слов равновесных кодов с отмеченным выше свойством. Например, равновесный $1/2$ -код содержит две комбинации $\{01, 10\}$ и используется при кодировании информации для обнаружения ошибок, к примеру, в пространственном парафазном представлении сигналов [25]. Равновесный $2/4$ -код содержит шесть комбинаций $\{0011, 0101, 0110, 1001, 1010, 1100\}$ и используется при организации схем встроенного контроля для комбинационных составляющих дискретных устройств [26].

В работе [24] предложено использовать равновесные коды вида $\frac{n}{2}/n$ при синтезе самопроверяемых дискретных устройств с контролем вычислений по двум диагностическим признакам – по принадлежности кодовых слов заданному равновесному коду и по принадлежности каждой функции, описывающей символы кодовых слов, к классу самодвойственных булевых функций. Для этого необходимо в процессе синтеза схемы встроенного контроля (СВК) обеспечить следующее свойство: на ортогональных по всем разрядам наборах значений аргументов требуется формировать в СВК ортогональные по всем разрядам кодовые слова равновесных кодов.

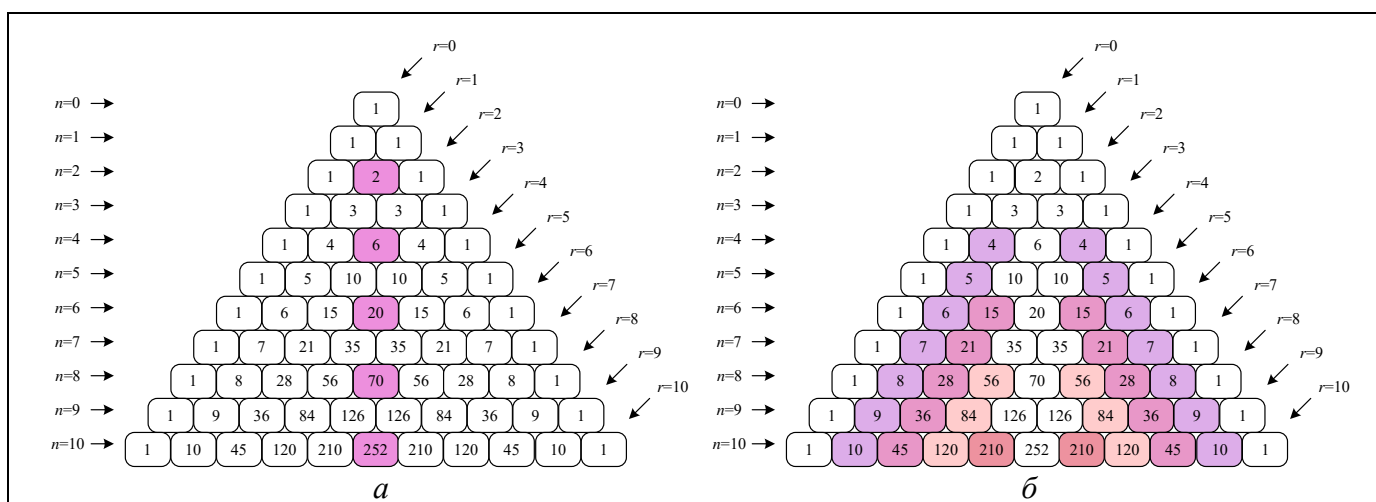


Рис. 1. Варианты выделения на треугольнике Паскаля а) одного и б) двух равновесных кодов с ортогональными по всем разрядам кодовыми словами

Исследования возможностей применения равновесных кодов при контроле вычислений на выходах дискретных устройств по двум диагностическим признакам привели к задаче определения возможностей применения групп иных равновесных кодов для организации СВК. Это позволило бы иметь большую вариативность в выборе способов синтеза СВК и влияния на ключевые показатели эффективности реализации самопроверяемых дискретных устройств.

Если объединить множества кодовых слов равновесных кодов r/n и $(n-r)/n$, где $r \in \left\{1, 2, \dots, \left\lfloor \frac{n}{2} \right\rfloor\right\}$ – вес кодового слова, то получим

композицию пары равновесных кодов, кодовые слова которой обладают искомым свойством. На треугольнике Паскаля, расположенном на рис. 1, б, соответствующие таким парам равновесных кодов биномиальные коэффициенты выделены одним цветом. Не выделены на рисунке коды с $r=0$, вырожденные пары, для которых при четном n вес равен $r = \frac{n}{2}$, а также пары кодов, для которых вес различается на единицу.

Утверждение 1. Композицией из пар кодов r/n и $(n-r)/n$ будут обнаруживаться любые одиночные ошибки в кодовых словах при выполнении следующего условия:

$$n - 2r \geq 2. \quad (1)$$

Доказательство. Вес кодовых слов кода r/n равен r , а вес кодовых слов кода $(n-r)/n$ равен $n-r$. Кодами r/n , также как и кодами $(n-r)/n$, будут обнаруживаться любые одиночные ошибки, переводящие кодовые слова, принадлежащие данному коду, в кодовые слова этого же кода. Не обнаруживаться могут только

одиночные ошибки, переводящие кодовые слова кода r/n в кодовые слова кода $(n-r)/n$. Это возможно только в том случае, если вес кодовых слов кода $(n-r)/n$ больше веса кода r/n на единицу. Иными словами, $(n-r) - r = 1$, или же $n - 2r = 1$. Отсюда становится ясной справедливость неравенства (1) для обнаружения любых одиночных ошибок. ♦

Таким образом, получается целая группа модифицированных кодов, основанных на использовании пар равновесных кодов. Их применение позволяет решать некоторые задачи синтеза контролепригодных и самопроверяемых дискретных устройств, а также их диагностического обеспечения.

2. ХАРАКТЕРИСТИКИ ОБНАРУЖЕНИЯ ОШИБОК КОМПОЗИЦИЯМИ ДВУХ РАВНОВЕСНЫХ КОДОВ

Определим, какое количество ошибок не будет обнаруживаться композициями двух равновесных кодов, а также дадим их характеристику.

Утверждение 2. Общее количество необнаруживаемых выбранной композицией из пар кодов r/n и $(n-r)/n$ ошибок в кодовых словах определяется по формуле

$$N_{ND} = 2C_n^r (2C_n^r - 1). \quad (2)$$

Доказательство. Мощность множества кодовых слов кода r/n равна C_n^r . Мощность множества кодовых слов кода $(n-r)/n$ равна $C_n^{n-r} = C_n^r$. Отсюда следует, что композиция образована $2C_n^r$ кодовыми словами. Ошибка не будет обнаружена тогда и только тогда, когда она переведет кодовое слово данной композиции кодов в кодовое слово, принадлежащее кодам из данной композиции. Всего таких переходов для каждо-

го кодового слова $2C_n^r - 1$. Отсюда становится очевидной справедливость выражения (2). ♦

Обозначим далее композицию пар равновесных кодов как $(r/n + (n - r)/n)$ -коды. Например, для первых значений n они приведены в табл. 1. Цветом в ней обозначены равновесные коды вида

$\frac{n}{2} / n$, также обладающие свойством присутствия

во множестве кодовых слов пар ортогональных по всем разрядам комбинаций.

В табл. 2 дается характеристика необнаруживаемых $(r/n + (n - r)/n)$ -кодами ошибок. Для каждого

Таблица 1

Композиции пар равновесных кодов

n	r					
	1	2	3	4	5	6
4	1/4+3/4	2/4	–	–	–	–
5	1/5+4/5	–	–	–	–	–
6	1/6+5/6	2/6+4/6	3/6	–	–	–
7	1/7+6/7	2/7+5/7	–	–	–	–
8	1/8+7/8	2/8+6/8	3/8+5/8	4/8	–	–
9	1/9+8/9	2/9+7/9	3/9+6/9	–	–	–
10	1/10+9/10	2/10+8/10	3/10+7/10	4/10+6/10	5/10	–
11	1/11+10/11	2/11+9/11	3/11+8/11	4/11+7/11	–	–
12	1/12+11/12	2/12+10/12	3/12+9/12	4/12+8/12	5/12+7/12	6/12
13	1/13+12/13	2/13+11/13	3/13+10/13	4/13+9/13	5/13+8/13	–
14	1/14+13/14	2/14+12/14	3/14+11/14	4/14+10/14	5/14+9/14	6/14+8/14
15	1/15+14/15	2/15+13/15	3/15+12/15	4/15+11/15	5/15+10/15	6/15+9/15

Таблица 2

Характеристика не обнаруживаемых $(r/n + (n - r)/n)$ -кодами ошибок

n	Число необнаруживаемых ошибок и их доля от общего количества ошибок в кодовых словах					
	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$
4	56 23,33333 %	30 12,5 %	–	–	–	–
5	90 9,07258 %	–	–	–	–	–
6	132 3,27381 %	870 21,57738 %	380 9,4246 %	–	–	–
7	182 1,11959 %	1722 10,59301 %	–	–	–	–
8	240 0,36765 %	3080 4,71814 %	12432 19,04412 %	4830 7,3989 %	–	–
9	306 0,11696 %	5112 1,95389 %	28056 10,72346 %	–	–	–
10	380 0,03628 %	8010 0,76464 %	57360 5,47562 %	175980 16,79917 %	63252 6,03808 %	–
11	462 0,01102 %	11990 0,286 %	108570 2,58978 %	434940 10,37484 %	–	–
12	552 0,00329 %	17292 0,10309 %	193160 1,1516 %	979110 5,83738 %	2507472 14,94935 %	852852 5,08464 %
13	650 0,00097 %	24180 0,03604 %	326612 0,48675 %	2043470 3,04538 %	6622902 9,8701 %	–
14	756 0,00028 %	32942 0,01227 %	529256 0,19718 %	4006002 1,49244 %	16028012 5,97126 %	36066030 13,43646 %
15	870 0,00008 %	43890 0,00409 %	827190 0,07704 %	7450170 0,69387 %	36066030 3,35901 %	100190090 9,33121 %

кода указаны два числа (сверху и снизу в графе) – число необнаруживаемых ошибок, а также доля необнаруживаемых ошибок от общего их количества в кодовых словах (γ_n , %). Композиции двух равновесных кодов обнаруживают существенное количество ошибок. С ростом значения n для каждого r доля необнаруживаемых ошибок постепенно снижается (рис. 2).

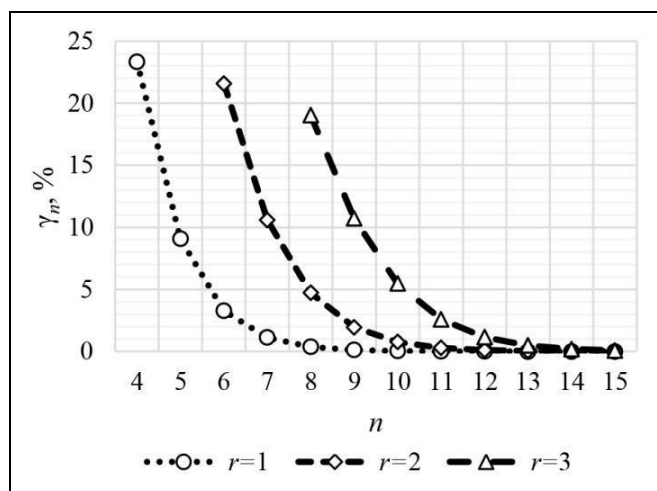


Рис. 2. Зависимости γ_n от n для различных $(r/n + (n-r)/n)$ -кодов

Дадим характеристику не обнаруживаемых различными $(r/n + (n-r)/n)$ -кодами ошибок по видам (монотонные, симметричные и асимметричные) и кратностям d .

Для начала рассмотрим коды вида $(1/n + (n-1)/n)$.

Возьмем первый код $(1/4 + 3/4)$. Выберем произвольное кодовое слово $1/4$ -кода, например, $\langle 0001 \rangle$. Оно может исказиться в кодовые слова, принадлежащие композиции данных кодов, $2C_4^1 - 1 = 7$ способами. Количество искажений кодового слова $\langle 0001 \rangle$ в кодовые слова, принадлежащие коду $1/4$, равно $C_4^1 - 1 = 3$. Все эти искажения в силу того, что сохраняется вес кодового слова, будут двукратными симметричными искажениями. Исказиться в кодовые слова, принадлежащие коду $3/4$, кодовое слово $\langle 0001 \rangle$ может $C_4^3 = 4$ способами. Причем если искажение произойдет так, что единственная единица в кодовом слове сохранит свое значение, то ошибка будет монотонной. Так как вес кодовых слов $3/4$ -кода равен $n - r = 3$, то возможно $C_{4-1}^{3-1} = C_3^2 = 3$ способа таких искажений. Кратность монотонных необнаруживаемых ошибок будет равна двум. Остается единственное слово $3/4$ -кода $\langle 1110 \rangle$, ортогональное по всем раз-

рядам по отношению к кодовому слову $\langle 0001 \rangle$. Последнее может перейти в первое только при четырехкратной асимметричной ошибке.

Аналогичное обстоятельство имеет место при рассмотрении и всех оставшихся кодовых слов $(1/4 + 3/4)$ -кодов.

Возьмем второй код $(1/5 + 4/5)$. Выберем произвольное кодовое слово $1/5$ -кода, например, $\langle 00001 \rangle$. Оно может исказиться в кодовые слова, принадлежащие композиции данных кодов, $2C_5^1 - 1 = 9$ способами. Количество искажений кодового слова $\langle 00001 \rangle$ в кодовые слова, принадлежащие коду $1/5$, равно $C_5^1 - 1 = 4$. Все эти искажения в силу того, что сохраняется вес кодового слова, будут двукратными симметричными искажениями. Исказиться в кодовые слова, принадлежащие коду $4/5$, кодовое слово $\langle 00001 \rangle$ может $C_5^4 = 5$ способами. Причем если искажение произойдет так, что единственная единица в кодовом слове сохранит свое значение, то ошибка будет монотонной. Так как вес кодовых слов $4/5$ -кода равен $n - r = 4$, то возможны $C_{5-1}^{4-1} = C_4^3 = 4$ способа таких искажений. Кратность монотонных необнаруживаемых ошибок будет равна трем. Остается единственное слово $4/5$ -кода $\langle 11110 \rangle$, ортогональное по всем разрядам по отношению к кодовому слову $\langle 00001 \rangle$. Последнее может перейти в первое только при пятикратной асимметричной ошибке. Аналогично рассматриваются и все оставшиеся кодовые слова $(1/5 + 4/5)$ -кодов.

Можно продолжить ряд рассуждений и прийти к следующим положениям, справедливым для кодов вида $(1/n + (n-1)/n)$:

- для каждого кодового слова существует $n - 1$ вариант перехода в кодовое слово, принадлежащее только этому коду, характеризующего двукратной симметричной ошибкой;
- для каждого кодового слова существует $n - 1$ вариант перехода в кодовое слово, принадлежащее второму коду в композиции, характеризующего $(n - 2r) = (n - 2)$ -кратной монотонной ошибкой;
- для каждого кодового слова существует один вариант перехода в кодовое слово, принадлежащее второму коду в композиции, характеризующего n -кратной асимметричной ошибкой.

Так как всего имеется $2C_n^1 = 2n$ кодовых слов, то каждое из полученных чисел по каждому виду и кратности ошибки следует умножить на это число, что даст общее количество необнаруживаемых ошибок по видам и кратностям.

Введем следующие обозначения. Будем общее количество необнаруживаемых ошибок представлять в виде формулы

$$N_{ND} = N_{v,d} v_d + N_{\sigma,d} \sigma_d + N_{\alpha,d} \alpha_d, \quad (3)$$

где $N_{v,d}$, $N_{\sigma,d}$, $N_{\alpha,d}$ – количество необнаруживаемых монотонных, симметричных и асимметричных ошибок с кратностью d ; символом v_d обозначим принадлежность числа необнаруживаемых ошибок к классу монотонных с кратностью d , σ_d – принадлежность необнаруживаемых ошибок к классу симметричных с кратностью d , α_d – принадлежность необнаруживаемых ошибок к классу асимметричных с кратностью d .

Используя выражение (3) и проведенные выше рассуждения, запишем формулу необнаруживаемых ошибок для кодов вида $(1/n + (n-1)/n)$:

$$\begin{aligned} N_{ND}^{1/n+(n-1)/n} &= 2n((n-1)v_{n-2} + (n-1)\sigma_2 + 1\alpha_n) = \\ &= 2n(n-1)(v_{n-2} + \sigma_2) + 2n\alpha_n. \end{aligned} \quad (4)$$

Например, для $(1/4 + 3/4)$ -кодов формула (4) дает следующий результат:

$$\begin{aligned} N_{ND}^{1/4+3/4} &= 2 \cdot 4 \cdot (4-1)(v_2 + \sigma_2) + \\ &+ 2 \cdot 4 \alpha_4 = 24v_2 + 24\sigma_2 + 8\alpha_4. \end{aligned}$$

Обобщим полученный результат и рассмотрим далее композицию кодов вида $(2/n + (n-2)/n)$.

Возьмем в качестве примера $(2/6 + 4/6)$ -код. Выберем произвольное кодовое слово $2/6$ -кода, например, $\langle 000011 \rangle$. Существует $2C_6^2 - 1 = 29$ переходов его в кодовые слова, принадлежащие этой композиции. Среди них $C_6^2 - 1 = 14$ переходов в кодовые слова $2/6$ -кода, вызванные двукратными симметричными ошибками. Также имеется $C_6^2 = 15$ переходов в кодовые слова $4/6$ -кода. Структура таких ошибок следующая. Имеется один переход в кодовое слово $4/6$ -кода, ортогональное по всем разрядам к рассматриваемому слову $\langle 111100 \rangle$, вызванный шестикратной асимметричной ошибкой. Имеется $C_{6-2}^2 = C_4^2 = 6$ переходов, при которых единицы в кодовом слове $\langle 000011 \rangle$ сохраняют свои позиции и искажаются два нуля. Такие переходы вызваны двукратными монотонными ошибками. Также имеются переходы, при которых одна единица сохраняет свое значение в кодовом слове $\langle 000011 \rangle$, одна искажается, а также искажаются еще три нулевых разряда. Число вариантов искажений единиц равно C_2^1 , число вариантов искажений нулей – C_4^3 . Итого

имеем $C_4^3 C_2^1 = 8$ четырехкратных асимметричных искажений.

Такая структура ошибок присуща всем кодовым словам в рассматриваемой композиции.

Нетрудно обобщить полученное для $(2/n + (n-2)/n)$ -кодов:

– для каждого кодового слова существует $C_n^2 - 1$ вариант перехода в кодовое слово, принадлежащее только этому коду, характеризуемого двукратной симметричной ошибкой;

– для каждого кодового слова существует C_{n-2}^2 вариант перехода в кодовое слово, принадлежащее второму коду в композиции, характеризуемого $(n-2r) = (n-4)$ -кратной монотонной ошибкой;

– для каждого кодового слова существует $2C_{n-2}^{n-3}$ вариант перехода в кодовое слово, принадлежащее второму коду в композиции, характеризуемого четырехкратной асимметричной ошибкой;

– для каждого кодового слова существует один вариант перехода в кодовое слово, принадлежащее второму коду в композиции, характеризуемого n -кратной асимметричной ошибкой.

Так как всего имеется $2C_n^2$ кодовых слов, то каждое из полученных чисел по каждому виду и кратности ошибки следует умножить на это число, что даст общее количество необнаруживаемых ошибок по видам и кратностям:

$$\begin{aligned} N_{ND}^{2/n+(n-2)/n} &= 2C_n^2 (C_{n-2}^2 v_2 + (C_n^2 - 1)\sigma_2 + \\ &+ 2C_{n-2}^{n-3} \alpha_4 + 1\alpha_n) = 2C_n^2 C_{n-2}^2 v_2 + \\ &+ 2C_n^2 (C_n^2 - 1)\sigma_2 + 4C_n^2 C_{n-2}^{n-3} \alpha_4 + 2C_n^2 \alpha_n. \end{aligned} \quad (5)$$

Например, для $(2/6 + 4/6)$ -кодов формула (5) дает следующий результат:

$$\begin{aligned} N_{ND}^{2/6+4/6} &= 2C_6^2 C_4^2 v_2 + 2C_6^2 (C_6^2 - 1)\sigma_2 + \\ &+ 4C_6^2 C_4^3 \alpha_4 + 2C_6^2 \alpha_6 = 180v_2 + \\ &+ 420\sigma_2 + 240\alpha_4 + 30\alpha_6. \end{aligned}$$

Для обобщения и рассмотрения $(r/n + (n-r)/n)$ -кодов характерно следующее:

– для каждого кодового слова существует $C_n^r - 1$ вариант перехода в кодовое слово, принадлежащее только этому коду, характеризуемого двукратной симметричной ошибкой;

– для каждого кодового слова существует C_{n-r}^{r-r} вариант перехода в кодовое слово, принадлежащее второму коду в композиции, характеризуемого $(n-2r)$ -кратной монотонной ошибкой;



– для каждого кодового слова существует

$$\begin{aligned}
 & C_r^1 C_{n-r}^{(n-r)-(r-1)} + C_r^2 C_{n-r}^{(n-r)-(r-2)} + \dots + \\
 & + C_r^{r-1} C_{n-r}^{n-r-(r-(r-1))} + C_r^r C_{n-r}^{n-r-(r-r)} = \\
 & = C_r^1 C_{n-r}^{n-2r+1} + C_r^2 C_{n-r}^{n-2r+2} + \dots + C_r^{r-1} C_{n-r}^{n-r-1} + \\
 & + C_r^r C_{n-r}^{n-r} = \sum_{i=1}^r C_r^i C_{n-r}^{(n-r)-(r-i)}
 \end{aligned}$$

вариантов перехода в кодовое слово, принадлежащее второму коду в композиции, характеризуемого асимметричными ошибками, причем первое слагаемое характеризует количество асимметричных ошибок кратности $n-2r+2$, второе – количество асимметричных ошибок кратности $n-2r+4$, ..., $r-1$ -е слагаемое – количество асимметричных ошибок кратности $((n-r)-(r-(r-1)) + (r-1)) = (n-2)$, r -е слагаемое – количество асимметричных ошибок кратности n .

Общее количество кодовых слов в рассматриваемой композиции равно $2C_n^r$.

Отсюда приходим к общей формуле, описывающей количество необнаруживаемых ошибок различных видов и кратностей $(r/n + (n-r)/n)$ -кодах:

$$\begin{aligned}
 N_{ND}^{r/n+(n-r)/n} &= 2C_n^r \left(C_{n-r}^r v_{n-2r} + (C_n^r - 1) \sigma_2 + \right. \\
 &+ C_r^1 C_{n-r}^{n-2r+1} \alpha_{n-2r+2} + C_r^2 C_{n-r}^{n-2r+2} \alpha_{n-2r+4} + \dots \\
 &+ C_r^{r-1} C_{n-r}^{n-r-1} \alpha_{n-2} + C_r^r C_{n-r}^{n-r} \alpha_n \left. \right) = \\
 &= 2C_n^r C_{n-r}^r v_{n-2r} + 2C_n^r (C_n^r - 1) \sigma_2 + \\
 &+ 2C_n^r C_r^1 C_{n-r}^{n-2r+1} \alpha_{n-2r+2} + 2C_n^r C_r^2 C_{n-r}^{n-2r+2} \alpha_{n-2r+4} + \dots \\
 &+ 2C_n^r C_r^{r-1} C_{n-r}^{n-r-1} \alpha_{n-2} + 2C_n^r C_r^r C_{n-r}^{n-r} \alpha_n.
 \end{aligned} \quad (6)$$

Формулы (4) и (5) являются частными случаями формулы (6).

Полный учет характеристик обнаружения ошибок в кодовых словах композиций равновесных кодов позволяет на практике выработать алгоритмы их использования при построении дискретных устройств с обнаружением неисправностей.

3. КОМПОЗИЦИИ ЧЕТНОГО ЧИСЛА РАВНОВЕСНЫХ КОДОВ

Для $n \geq 8$ могут быть построены композиции не двух, а большего числа равновесных кодов, которые будут обладать свойством содержания во множестве кодовых слов исключительно пар ортогональных по всем разрядам кодовых слов. Такие коды строятся при четных значениях $n \geq 8$. На рис. 3 на треугольнике Паскаля выделены четверки равновесных кодов, объединения множеств кодовых слов которых позволяет сформировать искомого композицию.

В табл. 3 приведен состав композиций четверок равновесных кодов, а в табл. 4 дана характеристика не обнаруживаемых ими ошибок. Композиции четверок равновесных кодов при $n > 8$ не обнаруживают сотые и тысячные доли процента от общего числа ошибок, возникающих в кодовых словах с соответствующей длиной кодовых слов.

Следуя за рассуждениями по определению характеристик обнаружения ошибок по видам и кратностям для пар композиций равновесных кодов, можно дать более детальную характеристику ошибок, не обнаруживаемых рассматриваемыми композициями четверок равновесных кодов.

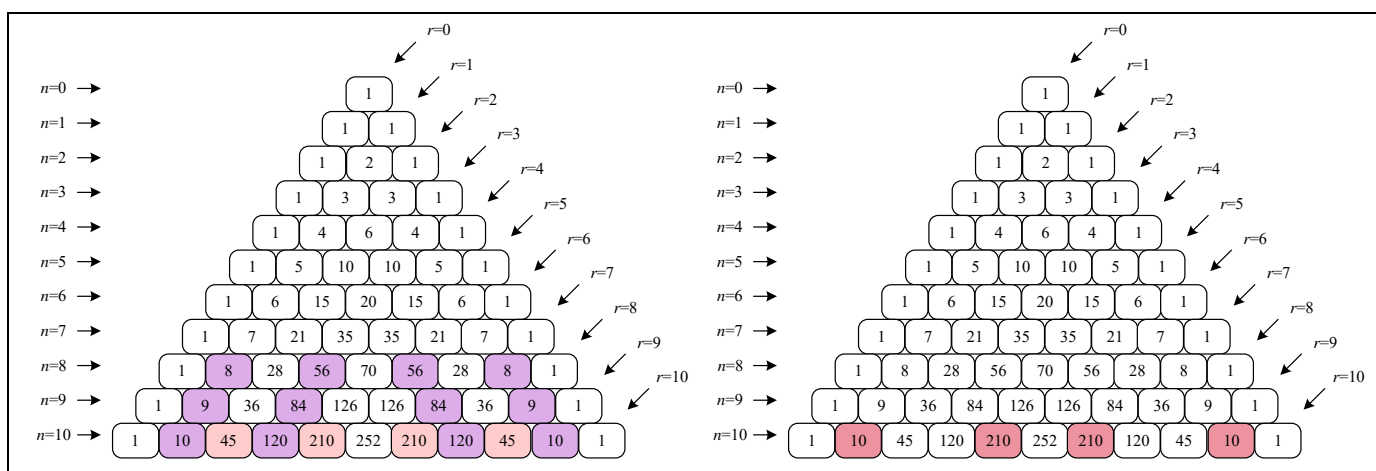


Рис. 3. Варианты выделения на треугольнике Паскаля четырех равновесных кодов с ортогональными по всем разрядам комбинациями

Таблица 3

Композиции четверок равновесных кодов

n	r_1, r_2									
	$r_1=1, r_2=3$	$r_1=1, r_2=4$	$r_1=1, r_2=5$	$r_1=1, r_2=6$	$r_1=2, r_2=4$	$r_1=2, r_2=5$	$r_1=2, r_2=6$	$r_1=3, r_2=5$	$r_1=3, r_2=6$	$r_1=4, r_2=6$
8	1/8+7/8, 3/8+5/8	–	–	–	–	–	–	–	–	–
9	1/9+8/9, 3/9+6/9	–	–	–	–	–	–	–	–	–
10	1/10+9/10, 3/10+7/10	1/10+9/10, 4/10+6/10	–	–	2/10+8/10, 4/10+6/10	–	–	–	–	–
11	1/11+10/11, 3/11+8/11	1/11+10/11, 4/11+7/11	–	–	2/11+9/11, 4/11+7/11	–	–	–	–	–
12	1/12+11/12, 3/12+9/12	1/12+11/12, 4/12+8/12	1/12+11/12, 5/12+7/12	–	2/12+10/12, 4/12+8/12	2/12+10/12, 5/12+7/12	–	3/12+9/12, 5/12+7/12	–	–
13	1/13+12/13, 3/13+10/13	1/13+12/13, 4/13+9/13	1/13+12/13, 5/13+8/13	–	2/13+11/13, 4/13+9/13	2/13+11/13, 5/13+8/13	–	3/13+10/13, 5/13+8/13	–	–
14	1/14+13/14, 3/14+11/14	1/14+13/14, 4/14+10/14	1/14+13/14, 5/14+9/14	1/14+13/14, 6/14+8/14	2/14+12/14, 4/14+10/14	2/14+12/14, 5/14+9/14	2/14+12/14, 6/14+8/14	3/14+11/14, 6/14+8/14	3/14+11/14, 6/14+8/14	4/14+10/14, 6/14+8/14
15	1/15+14/15, 3/15+12/15	1/15+14/15, 4/15+11/15	1/15+14/15, 5/15+10/15	1/15+14/15, 6/15+9/15	2/15+13/15, 4/15+11/15	2/15+13/15, 5/15+10/15	2/15+13/15, 6/15+9/15	3/15+12/15, 6/15+9/15	3/15+12/15, 6/15+9/15	4/15+11/15, 6/15+9/15

Таблица 4

Характеристика ошибок, не обнаруживаемых композициями из четырех равновесных кодов

n	Число необнаруживаемых ошибок и их доля от общего количества ошибок в кодовых словах									
	$r_1=1, r_2=3$	$r_1=1, r_2=4$	$r_1=1, r_2=5$	$r_1=1, r_2=6$	$r_1=2, r_2=4$	$r_1=2, r_2=5$	$r_1=2, r_2=6$	$r_1=3, r_2=5$	$r_1=3, r_2=6$	$r_1=4, r_2=6$
8	128 0,19608 %	–	–	–	–	–	–	–	–	–
9	186 0,07109 %	–	–	–	–	–	–	–	–	–
10	260 0,02482 %	440 0,042 %	–	–	510 0,04868 %	–	–	–	–	–
11	352 0,0084 %	682 0,01627 %	–	–	770 0,01837 %	–	–	–	–	–
12	464 0,00277 %	1014 0,00605 %	1608 0,00959 %	–	1122 0,00669 %	1716 0,01023 %	–	2024 0,01207 %	–	–
13	598 0,00089 %	1456 0,00217 %	2600 0,00387 %	–	1586 0,00236 %	2730 0,00407 %	–	3146 0,00469 %	–	–
14	756 0,00028 %	2030 0,00076 %	4032 0,0015 %	6034 0,00225 %	2184 0,00081 %	4186 0,00156 %	6188 0,00231 %	4732 0,00176 %	6734 0,00251 %	8008 0,00298 %
15	940 0,00009 %	2760 0,00026 %	6036 0,00056 %	10040 0,00094 %	2940 0,00027 %	6216 0,00058 %	10220 0,00095 %	6916 0,00064 %	10920 0,00102 %	12740 0,00119 %

4. ПРИМЕНЕНИЕ КОМПОЗИЦИЙ ДВУХ РАВНОВЕСНЫХ КОДОВ В ЗАДАЧАХ СИНТЕЗА САМОПРОВЕРЯЕМЫХ ДИСКРЕТНЫХ УСТРОЙСТВ

На рис. 4 изображена структура организации самопроверяемого дискретного устройства со схемой встроенного контроля, реализованной на основе $(r/n + (n-r)/n)$ -кода. Она позволяет учесть два диагностических признака при контроле вычислений: принадлежность формируемых кодовых слов заданному $(r/n + (n-r)/n)$ -коду и принадлежность каждой функции, описывающей проверочные символы, к классу самодвойственных функций.

Объектом диагностирования является комбинационное дискретное устройство $F(X)$, вычисляющее значения булевых функций $f_1(X), f_2(X), \dots, f_{n-1}(X), f_n(X)$ при подаче на входы наборов значений аргументов $\langle x_t \ x_{t-1} \dots x_2 \ x_1 \rangle = \langle X \rangle$. Для контроля вычислений устройство $F(X)$ дополняется специализированной СВК. Кроме того, работа схемы настраивается специальным образом [27].

Выходы устройства $F(X)$ подключены к входам блока коррекции сигналов (БКС) в СВК. Этот блок предназначен для преобразования сигналов, формируемых на выходах устройства $F(X)$, в сигналы $h_1(X), h_2(X), \dots, h_{n-1}(X), h_n(X)$ на каждом наборе значений аргументов. Блок коррекции сигналов образован каскадом двухвходовых элементов сложения по модулю $M=2$ (XOR), на первые входы каждого из которых подаются сигналы $f_1(X), f_2(X), \dots, f_{n-1}(X), f_n(X)$, а на вторые – сигналы $g_1(X), g_2(X), \dots, g_{n-1}(X), g_n(X)$ с такими же нижними индексами,

формируемые блоком $G(X)$ вычисления функций логической коррекции:

$$h_i(X) = f_i(X) \oplus g_i(X), i = \overline{1, n}. \quad (7)$$

Преобразование (7) осуществляется на каждом наборе значений аргументов и позволяет наделить функции $h_1(X), h_2(X), \dots, h_{n-1}(X), h_n(X)$ следующими двумя свойствами. Первое состоит в том, что кодовое слово $\langle h_n(X) \ h_{n-1}(X) \dots h_2(X) \ h_1(X) \rangle$ должно принадлежать выбранному $(r/n + (n-r)/n)$ -коду. Второе свойство заключается в том, что значения каждой функции $h_1(X), h_2(X), \dots, h_{n-1}(X), h_n(X)$ на каждом наборе значений аргументов формируются таким образом, чтобы сами функции оказались самодвойственными. Как отмечалось выше, это требует только того, чтобы на ортогональных наборах значений аргументов формировались ортогональные по всем разрядам комбинации, принадлежащие $(r/n + (n-r)/n)$ -коду. Необходимо доопределение значений векторов $\langle h_n(X) \ h_{n-1}(X) \dots h_2(X) \ h_1(X) \rangle$ на всех наборах значений аргументов таким образом, чтобы хотя бы единожды были сформированы все кодовые слова, принадлежащие r/n -коду и $(n-r)/n$ -коду. Эта особенность требует выполнения ограничения

$$t \geq \lceil \log_2 2C_n^r \rceil.$$

Контроль принадлежности кодовых слов $\langle h_n(X) \ h_{n-1}(X) \dots h_2(X) \ h_1(X) \rangle$ выбранному $(r/n + (n-r)/n)$ -коду осуществляется с помощью самопроверяемого тестера данного кода – устройства TSC (*totally self-checking checker*). Такие тестеры могут синтезироваться различными способами, например, с выделением двух функций

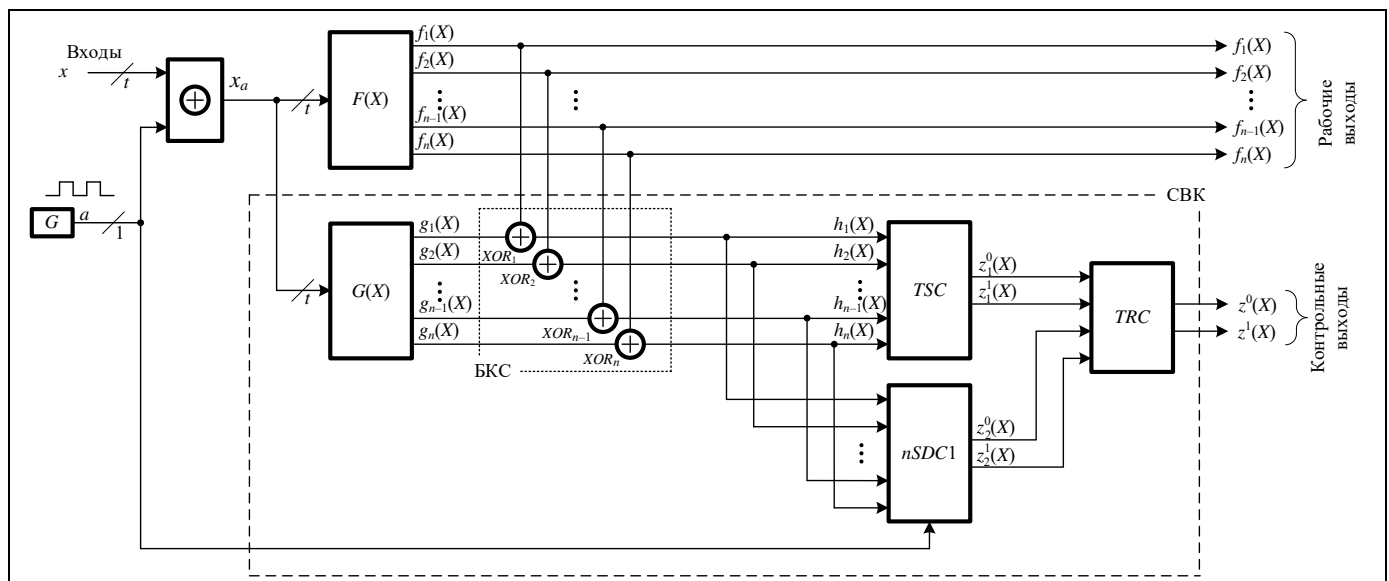


Рис. 4. Структура организации самопроверяемого дискретного устройства

$z_1^0(X)$ и $z_1^1(X)$, каждая из которых является дизъюнкцией конъюнкций, соответствующих кодовым словам r/n и $(n-r)/n$ кода (они, к слову, могут входить в каждую из функций, но только единожды!). Можно использовать скобочные формы для упрощения структур тестеров. Методы же синтеза тестеров равновесных кодов, изложенные в работах [9, 10], напрямую не применимы для синтеза тестеров их композиций.

Контроль принадлежности каждой функции $h_1(X)$, $h_2(X)$, ..., $h_{n-1}(X)$, $h_n(X)$ к классу самодвойственных булевых функций осуществляется с использованием устройства $nSDC1$, на входы которого поступают n самодвойственных сигналов, а на выходах формируются значения двух функций $z_2^0(X)$ и $z_2^1(X)$. Такое устройство может быть реализовано двумя способами. Первый из них предполагает установку n элементарных самопроверяемых тестеров самодвойственных сигналов SDC (*self-dual checker*) [27] и $n-1$ элементарного модуля сжатия парафазных сигналов TRC (*two-rail checker*) [28]. Второй способ основан на использовании предварительного сжатия n самодвойственных сигналов в один с использованием трехвходовых элементов XOR по схеме из работы [29] и контроль полученного сигнала с использованием одного устройства SDC .

Выходы $z_1^0(X)$ и $z_1^1(X)$, а также $z_2^0(X)$ и $z_2^1(X)$ функционируют в парафазной логике. Нарушение парафазности свидетельствует о наличии ошибки в вычислениях и косвенно говорит о наличии неисправностей в одном из блоков. Выходы $z_1^0(X)$ и $z_1^1(X)$, а также $z_2^0(X)$ и $z_2^1(X)$ подключаются ко входам одного модуля TRC для получения единственного контрольного сигнала $Z^0(X)$ и $Z^1(X)$.

Принципы организации контроля самодвойственных сигналов требуют последовательного сравнения значения на прямой подаваемой комбинации и на ортогональной по всем разрядам, что требует снижения быстродействия вдвое [30]. Тогда первая подаваемая комбинация является рабочей, а вторая – контрольной. Таким образом, требуется подача на входы устройства пар наборов значений аргументов, составленных из рабочих комбинаций и ортогональных по всем переменным к ним контрольных комбинаций. Удобным образом это организуется при импульсном режиме работы структуры, приведенной на рис. 4. Для такого режима работы устанавливается генератор G прямоугольных импульсов со скважностью $S=2$, на выходе которого формируется последовательность

a 0101...01. Она подается на вторые входы двухвходовых элементов XOR , устанавливаемых для каждого входа $x_1, x_2, \dots, x_{t-1}, x_t$. На первые входы элементов XOR подаются значения самих входных переменных. Такая структура схемы позволяет преобразовать сигнал логического нуля в последовательность 0101...01, а сигнал логической единицы – в последовательность 1010...10. Таким образом, наборы значений аргументов поступают на входы блока $F(X)$ парами – рабочий и контрольный, ортогональный по всем разрядам к рабочему. Также сигнал от генератора прямоугольных импульсов необходим для работы тестеров самодвойственных сигналов SDC в СВК.

Отдельные примеры синтеза СВК по приведенному здесь методу оставим за рамками настоящей статьи. Отметим лишь, что использование при синтезе СВК по представленной на рис. 4 структуре $(r/n + (n-r)/n)$ -кодов позволяет на практике организовывать полностью самопроверяемые дискретные устройства с различными показателями эффективности. Прежде всего, здесь важны покрытие ошибок на выходах устройства $F(X)$, достижение свойства самопроверяемости СВК, а также получение наиболее простой реализации СВК, позволяющей синтезировать менее избыточные структуры, чем при использовании метода дублирования. Учитывая высокую вариативность в получении значений функций $g_1(X)$, $g_2(X)$, ..., $g_{n-1}(X)$, $g_n(X)$ на каждом наборе значений аргументов, можно формировать значительное их количество и выбирать среди них наиболее простые по сложности. Этот показатель определяет сложность реализуемого устройства [31]. Так, количество способов доопределения значений функций существенно зависит от числа t входов объекта диагностирования $F(X)$. Достаточно доопределить значения функций на первой половине наборов, число которых равно 2^{t-1} . На каждом из данных наборов можно осуществить доопределение $2C_n^r$ способами. На второй половине наборов значений аргументов значения функций получаются, исходя из свойств самодвойственных функций: требуется на ортогональных по всем разрядам наборах доопределить функции до противоположных значений. Итого имеется $2^t C_n^r$ способов доопределения значений функций, реализуемых на выходах блока $G(X)$. Имея большое количество способов построения блоков $G(X)$, можно выбирать наиболее простые их реализации, обеспечивая при этом и тестируемость компонентов СВК. Во многих практических случаях можно добиться более простой реализации блока $G(X)$, чем блока $F(X)$, что в су-



существенной мере влияет на аппаратную избыточность СВК и позволяет строить самопроверяемые устройства, менее избыточные в сравнении с использованием дублирования, обладающие к тому же улучшенными показателями контролепригодности [23]. Тем не менее, более простую реализацию, чем при использовании дублирования, применяя рассматриваемую структуру, построить удастся не для любых блоков $F(X)$, что является недостатком применения СВК с коррекцией всех сигналов от объекта диагностирования [32]. Исследования показывают, что при определенных условиях можно сокращать количество выходов блока $G(X)$ путем использования одинаковых функций для коррекции различных сигналов от блока $F(X)$. Такой подход работает только в частных случаях и до конца не исследован, но также может рассматриваться с точки зрения перспективы совершенствования алгоритмов синтеза СВК по структуре, приведенной на рис. 4.

ЗАКЛЮЧЕНИЕ

Использование композиций двух равновесных кодов r/n и $(n-r)/n$ при $r \in \left\{1, 2, \dots, \left\lfloor \frac{n}{2} \right\rfloor\right\}$ с учетом выполнения условия (1) позволяет существенно расширить число способов организации контролепригодных и самопроверяемых дискретных устройств с контролем вычислений по двум диагностическим признакам по сравнению с использованием в этих целях исключительно кодов $\frac{n}{2}/n$ с четными значениями n . Композиции равновесных кодов с ортогональными по всем разрядам комбинациями могут быть выделены при $n \geq 4$. С ростом значения n растет и число кодовых слов в таких кодах, а число способов выделения именно двух равновесных кодов увеличивается всякий раз на единицу для каждого четного n , начиная с $n = 4$. При $n \geq 8$ возможно построение композиции четного числа равновесных кодов с ортогональными по всем разрядам комбинациями.

Применение широкого класса композиций равновесных кодов с ортогональными по всем разрядам комбинациями позволяет как подбирать композицию под заданное число выходов объекта диагностирования, так и выбирать композицию с учетом возможного выделения групп выходов с меньшим их количеством для организации отдельных подсистем контроля вычислений. Интерес может вызывать применение композиций равновесных кодов с малыми значениями n , так как для них

будут более простыми тестеры и условия обеспечения их полной самопроверяемости.

Дальнейшие исследования могут быть связаны с разработкой методов синтеза СВК на основе композиций равновесных кодов с учетом минимизации различных показателей – структурной избыточности, контролепригодности, энергопотребления и т. д. Интересными могут оказаться практические приложения данных кодов, в особенности в тех областях науки и техники, где входные данные меняются редко (такие системы распространены на практике – к ним можно отнести комплексы противовоздушной обороны, системы электрической централизации стрелок и сигналов на железных дорогах, управляющие комплексы в атомной промышленности и др.). В них особо актуальными являются задачи рабочего диагностирования и обеспечения свойства готовности устройств [33, 34].

Описанные в настоящей статье принципы выделения композиций равновесных кодов, а также их свойства могут оказаться востребованными при построении высоконадежных дискретных систем на современной и перспективной элементной базе [35, 36].

ЛИТЕРАТУРА

1. Freiman, C.V. Optimal Error Detection Codes for Completely Asymmetric Binary Channels // Information and Control. – 1962. – Vol. 5, iss. 1. – P. 64–71. – DOI: 10.1016/S0019-9958(62)90223-1
2. Tallini, L.G., Bose, B. Design of Balanced and Constant Weight Codes for VLSI systems // IEEE Transactions on Computers. – 1998. – Vol. 47, no. 5. – P. 556–572. – DOI: 10.1109/12.677239
3. Butler, J.T., Sasao, T. Index to Constant Weight Codeword Converter // Proceedings of 2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS). – Seoul, South Korea, 2011. – DOI: 10.1109/MWSCAS.2011.6026312
4. Самофалов К.Г., Романкевич А.М., Валуцкий В.Н. и др. Прикладная теория цифровых автоматов / Под ред. К.Г. Самофалова. – Киев: «Вища школа», 1987. – 375 с. [Samofalov, K.G., Romankevich, A.M., Valujskij, V.N., et al. Prikladnaya teoriya cifrovyyh avtomatov / Pod red. K.G. Samofalova. – Kiev: «Vishcha shkola», 1987. – 375 s. (In Russian)].
5. Согомонян Е.С., Слабаков Е.В. Самопроверяемые устройства и отказоустойчивые системы. – М.: Радио и связь, 1989. – 208 с. [Sogomonyan, E.S., Slabakov, E.V. Samoproveryaemye ustroystva i otkazoustoichivye sistemy. – M.: Radio i svyaz', 1989. – 208 s. (In Russian)].
6. Gössel, M., Ocheretny, V., Sogomonyan, E., Marienfeld, D. New Methods of Concurrent Checking: Edition 1. – Dordrecht: Springer Science+Business Media B.V., 2008. – 184 p.
7. Das, D.K., Roy, S.S., Dmitriyev, A., et al. Constraint Don't Cares for Optimizing Designs for Concurrent Checking by 1-

- out-of-3 Codes // Proceedings of the 10th International Workshops on Boolean Problems. – Freiberg, Germany, 2012. – P. 33–40.
8. Сапожников В.В., Сапожников Вл.В. Самопроверяемые тестеры для равновесных кодов // Автоматика и телемеханика. – 1992. – № 3. – С. 3–35. [Sapozhnikov, V.V., Sapozhnikov, V.I. Samoproveryaemye testery dlya ravnovesnykh kodov // Avtomatika i Telemekhanika. – 1992. – No. 3. – P. 3–35. (In Russian)].
 9. Сапожников В.В., Сапожников Вл.В. Самопроверяемые дискретные устройства. – СПб: Энергоатомиздат, 1992. – 224 с. [Sapozhnikov, V.V., Sapozhnikov, V.I. Samoproveryaemye diskretnye ustroystva. – SPb: Energoatomizdat, 1992. – 224 s. (In Russian)].
 10. Piestrak, S.J. Design of Self-Testing Checkers for Unidirectional Error Detecting Codes. – Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 1995. – 111 p.
 11. Tarnik, M. Design of Embedded Constant Weight Code Checkers Based on Averaging Operations // Proceedings of 2010 IEEE 16th International On-Line Testing Symposium. – Corfu, Greece, 2010. – DOI: 10.1109/IOLTS.2010.5560193
 12. Матросова А.Ю., Буторина Н.Б., Якимова Н.О. Синтез детекторов равновесных кодов с использованием монотонных функций // Известия высших учебных заведений. Физика. – 2013. – Т. 56, № 9-2. – С. 171–173. [Matrosova, A.Yu., Butorina, N.B., Yakimova, N.O. Sintez detektorov ravnovesnykh kodov s ispol'zovaniem monotonnykh funktsiy // Izvestiya vysshikh uchebnykh zavedenij. Fizika. – 2013. – Vol. 56, no. 9-2. – P. 171–173. (In Russian)].
 13. Ефанов Д.В. Некоторые особенности обнаружения ошибок равномерными неразделимыми кодами // Известия высших учебных заведений. Приборостроение. – 2019. – Т. 62, № 7. – С. 621–631. – DOI: 10.17586/0021-3454-2019-62-7-621-631. [Efanov, D.V. Nekotorye osobennosti obnaruzheniya oshibok ravnomernymi nerazdelimymi kodami // Izvestiya vysshikh uchebnykh zavedenij. Priborostroenie. – 2019. – Vol. 62, no. 7. – P. 621–631. – DOI: 10.17586/0021-3454-2019-62-7-621-631. (In Russian)].
 14. Кориунов А.Д. Некоторые нерешенные задачи дискретной математики и математической кибернетики // Успехи математических наук. – 2009. – Т. 64, № 5. – С. 3–20. – DOI: 10.4213/rm9322. [Korshunov, A.D. Nekotorye nereshennyye zadachi diskretnoy matematiki i matematicheskoy kibernetiki // Uspekhi matematicheskikh nauk. – 2009. – Vol. 64, no. 5. – P. 3–20. – DOI: 10.4213/rm9322. (In Russian)].
 15. Можаров Г.П. Отказоустойчивые компьютерные сети, построенные на основе комбинаторных блок-дизайнов // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. – 2016. – № 6. – С. 41–53. – DOI: 10.18698/0236-3933-2016-6-41-53. [Mozharov G.P. Otkazoustojchivye komp'yuternye seti, postroennyye na osnove kombinatornykh blok-dizajnov // Vestnik MG TU im. N.E. Bauman. Ser. Priborostroenie. – 2016. – No. 6. – P. 41–53. – DOI: 10.18698/0236-3933-2016-6-41-53. (In Russian)].
 16. Chioktour, V., Kakarountas, A. Adaptive BIST for Concurrent On-Line Testing on Combinational Circuits // Electronics. – 2022. – Vol. 19. – Art. no. 3193. – DOI: 10.3390/electronics11193193
 17. Sahana A.R., Chiraag V., Suresh G., et al. Application of Error Detection and Correction Techniques to Self-Checking VLSI Systems: An Overview // Proceedings of 2023 IEEE Guwahati Subsection Conference (GCON). – Guwahati, 2023. – DOI: 10.1109/GCON58516.2023.10183449
 18. Слабаков Е.В., Согомонян Е.С. Построение полностью самопроверяемых комбинационных устройств с использованием равновесных кодов // Автоматика и телемеханика. – 1980. – № 9. – С. 173–181. [Slabakov, E.V., Sogomonyan, E.S. Postroenie polnost'yu samoproveryaemykh kombinatsionnykh ustrojstv s ispol'zovaniem ravnovesnykh kodov // Avtomatika i Telemekhanika. – 1980. – No. 9. – P. 173–181 (In Russian)].
 19. Sapozhnikov, V.V., Morosov, A., Sapozhnikov, V.I., Göessel, M. A New Design Method for Self-Checking Unidirectional Combinational Circuits // Journal of Electronic Testing: Theory and Applications. – 1998. – Vol. 12, iss. 1-2. – P. 41–53. – DOI: 10.1023/A:1008257118423
 20. Ефанов Д.В., Сапожников В.В., Сапожников Вл.В. Условия обнаружения неисправности логического элемента в комбинационном устройстве при функциональном контроле на основе кода Бергера // Автоматика и телемеханика. – 2017. – № 5. – С. 152–165. [Efanov, D.V., Sapozhnikov, V.V., Sapozhnikov, V.I. Conditions for detecting a logical element fault in a combination device under concurrent checking based on Berger's code // Automation and Remote Control. – 2017. – Vol. 78, no. 5. – P. 891–901.]
 21. Ефанов Д.В., Сапожников В.В., Сапожников Вл.В. Синтез самопроверяемых комбинационных устройств на основе выделения специальных групп выходов // Автоматика и телемеханика. – 2018. – № 9. – С. 79–94. [Efanov D.V., Sapozhnikov V.V., Sapozhnikov V.I. Sintez samoproveryaemykh kombinatsionnykh ustrojstv na osnove vydeleniya special'nykh grupp vyhodov // Avtomation and Remote Control. – 2018. – No. 9. – P. 79–94. (In Russian)].
 22. Efanov, D.V., Sapozhnikov, V.V., Sapozhnikov, V.I. Organization of a Fully Self-Checking Structure of a Combinational Device Based on Searching for Groups of Symmetrically Independent Outputs // Automatic Control and Computer Sciences. – 2020. – Vol. 54, iss. 4. – P. 279–290. – DOI: 10.3103/S0146411620040045.
 23. Ефанов Д.В., Погодина Т.С. Исследование свойств самодвойственных комбинационных устройств с контролем вычислений на основе кодов Хэмминга // Информатика и автоматизация. – 2023. – Т. 22, № 2. – С. 349–392. – DOI: 10.15622/ia.22.2.5. [Efanov D.V., Pogodina T.S. Properties investigation of self-dual combinational devices with calculation control based on Hamming codes // Informatics and Automation. – 2023. – Vol. 22, no. 2. – P. 349–392. – DOI: 10.15622/ia.22.2.5]
 24. Efanov, D., Sapozhnikov, V., Sapozhnikov, V.I., et al. Self-Dual Complement Method up to Constant-Weight Codes for Arrangement of Combinational Logical Circuits Concurrent Error-Detection Systems // Proceedings of 17th IEEE East-West Design & Test Symposium (EWDTS'2019). – Batumi, Georgia, 2019. – P. 136–143. – DOI: 10.1109/EWDTS.2019.8884398.
 25. Lala, P.K. Self-Checking and Fault-Tolerant Digital Design. – San Francisco: Morgan Kaufmann Publishers, 2001. – 216 p.
 26. Сапожников В.В., Сапожников Вл.В., Ефанов Д.В. Построение самопроверяемых структур систем функционального контроля на основе равновесного кода «2 из 4» // Проблемы управления. – 2017. – № 1. – С. 57–64. [Sapozhnikov, V.V., Sapozhnikov, V.I., Efanov, D.V. Design of Self-Checking Concurrent Error Detection Systems Based on «2-



- out-of-4» Constant-Weight Code // Control Sciences. – 2017. – No. 1. – P. 57–64. (In Russian)].
27. Ефанов Д.В., Пивоваров Д.В. Тестеры самодвойственных и «близких» к ним сигналов // Известия высших учебных заведений. Приборостроение. – 2024. – Т. 67, № 1. – С. 5–19. – DOI: 10.17586/0021-3454-2024-67-1-5-19. [Efanov, D.V., Pivovarov, D.V. Testery samodvojstvennyh i «blizkih» k nim signalov // Izvestiya vysshih uchebnyh zavedenij. Priborostroenie. – 2024. – Vol. 67, no. 1. – P. 5–19. – DOI: 10.17586/0021-3454-2024-67-1-5-19 (In Russian)].
 28. Nikolos, D. Self-Testing Embedded Two-Rail Checkers // Journal of Electronic Testing: Theory and Applications. – 1998. – Vol. 12, iss. 1-2. – P. 69–79. – DOI: 10.1023/A:1008281822966
 29. Dmitriev, A., Sapozhnikov, V., Sapozhnikov, V., Goessel, M. New Self-Dual Circuits for Error Detection and Testing // VLSI Design. – 2000. – Vol. 11, iss. 1. – P. 1–21. – DOI: 10.1155/2000/84720
 30. Гессель М., Мошанин В.И., Сапожников В.В., Сапожников Вл.В. Обнаружение неисправностей в самопроверяемых комбинационных схемах с использованием свойств самодвойственных функций // Автоматика и телемеханика. – 1997. – № 12. – С. 193–200. [Gessel', M., Moshanin, V.I., Sapozhnikov, V.V., Sapozhnikov, V.V. Obnaruzhenie neispravnostej v samoproveryaemyh kombinacionnyh skhemah s ispol'zovaniem svojstv samodvojstvennyh funkcij // Avtomatika i Telenekhanika. – 1997. – No. 12. – P. 193–200. (In Russian)].
 31. Коршунов А.Д. Сложность вычислений булевых функций // Успехи математических наук. – 2012. – Т. 67, № 1 (403). – С. 97–168. – DOI: 10.4213/rm9459. [Korshunov, A.D. Slozhnost' vychislenij bulevykh funkcij // Uspekhi matematicheskikh nauk. – 2012. – Vol. 67, no. 1 (403). – P. 97–168. – DOI: 10.4213/rm9459. (In Russian)]
 32. Ефанов Д.В. Синтез самопроверяемых комбинационных устройств на основе метода логической коррекции сигналов с применением кодов Боуза – Лина // Информационные технологии. – 2023. – Т. 29, № 10. – С. 503–511. – DOI: 10.17587/it.29.503-511. [Efanov, D.V. Sintez samoproveryaemyh kombinacionnyh ustrojstv na osnove metoda logicheskoy korrekcii signalov s primeneniem kodov Bouza – Lina // Informacionnye tekhnologii. – 2023. – Vol. 29, no. 10. – P. 503–511. – DOI: 10.17587/it.29.503-511. (In Russian)].
 33. Drozd, A., Kharchenko, V., Antoshchuk, S., et al. Checkability of the Digital Components in Safety-Critical Systems: Problems and Solutions // Proceedings of 9th IEEE East-West Design & Test Symposium (EWDTS'2011). – Sevastopol, Ukraine, 2011. – P. 411–416. – DOI: 10.1109/EWDTS.2011.6116606.
 34. Сапожников Вл.В. Синтез систем управления движением поездов на железнодорожных станциях с исключением опасных отказов. – М.: Наука, 2021. – 229 с. [Sapozhnikov, V.V. Sintez sistem upravleniya dvizheniem poezdov na zheleznodorozhnyh stanciyah s isklyucheniem opasnyh отказов. – М.: Наука, 2021. – 229 с. (In Russian)]
 35. Hahanov, V., Litvinova, E., Davitadze, Z., et al. Truth Table Based Intelligent Computing // Proceedings of 31st International Conference “Mixed Design of Integrated Circuits and Systems”. – Gdansk, Poland, 2024. – DOI: 10.23919/MIXDES62605.2024.10614035
 36. Meliqyan, V., Davtyan, V., Harutyunyan, A., et al. Novel Adaptive Crossover Mechanism for Genetic Algorithm on Integrated Circuit Cell Placement // Proceedings of IEEE East-West Design & Test Symposium (EWDTS). – Yerevan, Armenia, 2024. – DOI: 10.1109/EWDTS63723.2024.10873659
- Статья представлена к публикации членом редколлегии В.Г. Лебедевым.
- Поступила в редакцию 10.03.2025,
после доработки 12.05.2025.
Принята к публикации 18.06.2025.
- Ефанов Дмитрий Викторович** – д-р техн. наук, Санкт-Петербургский политехнический университет Петра Великого, г. Санкт-Петербург; Российский университет транспорта (МИИТ), г. Москва,
✉ TrES-4b@yandex.ru
ORCID iD: <https://orcid.org/0000-0002-4563-6411>.
- © 2025 г. Ефанов Д. В.



Эта статья доступна по [лицензии Creative Commons «Attribution» \(«Атрибуция»\) 4.0 Всемирная](https://creativecommons.org/licenses/by/4.0/).

COMPOSITIONS OF TWO CONSTANT-WEIGHT CODES WITH ORTHOGONAL COMBINATIONS OVER ALL BITS FOR SELF-CHECKING DISCRETE DEVICE DESIGN

D. V. Efanov

Peter the Great Saint Petersburg Polytechnic University, St. Petersburg, Russia
Russian University of Transport, Moscow, Russia

✉ TrES-4b@yandex.ru

Abstract. This paper proposes using compositions of two constant-weight codes with orthogonal combinations over all bits in the design of controllable and self-checking discrete devices. With such codes, computation control at the outputs of discrete devices can be implemented via the attribute of belonging of the codewords to a given constant-weight code and, moreover, via the attribute of belonging of each function describing the codeword bits to the class of self-dual Boolean functions. It is shown how to construct noninterference codes based on the composition of two constant-weight codes with orthogonal combinations over all bits. Explicit formulas are derived to determine the number of errors undetectable by compositions of constant-weight codes by their types (by the number of erroneous 0s and 1s in codewords) and multiplicities. The properties of the codes under consideration are briefly described. The structure of concurrent error-detection circuits is presented for discrete devices based on the composition of two constant-weight codes with orthogonal combinations over all bits and computation control via two diagnostic attributes. The use of such compositions can be effective in building highly reliable discrete devices on various components.

Keywords: controllable and self-checking devices; computation control via two diagnostic attributes; compositions of constant-weight codes; error detection at discrete device outputs.



МАТЕМАТИЧЕСКАЯ МОДЕЛЬ АДАПТИВНОГО УПРАВЛЕНИЯ ТРАФИКОМ В ПОДВИЖНЫХ СЕТЯХ С ПЕРЕМЕННЫМ КАЧЕСТВОМ СИГНАЛА

А. А. Есин

АО «МирВайФай, г. Москва

✉ anton.esin@imm.am

Аннотация. Рассматривается задача математического моделирования процесса передачи трафика с применением методов агрегации в подвижных сетях в условиях высокой мобильности абонентов и пространственно неоднородного покрытия, включая зоны деградации сигнала (так называемые мёртвые зоны). Для повышения надёжности и устойчивости передачи данных используется агрегация трафика на уровне канала. Для исследования эффективности алгоритмов агрегации и их адаптации к параметрам работы сети и скорости перемещения абонентов предложена марковская модель канала связи. Модель основана на периодическом аффинном движении мобильного устройства между базовыми станциями, равномерно расположенными на прямой. В рамках модели вводятся понятия зоны устойчивого покрытия и переходной зоны, определяемые расстоянием до ближайшей и следующей базовой станции. Состояние канала описывается марковской цепью с состояниями, соответствующими дискретизации уровня сигнала: стабильное соединение, ухудшенное соединение и разрыв; переходы между состояниями моделируются марковским процессом с постоянными интенсивностями, параметры которого определяются на основе данных реальных сетей. Рассмотрено также обобщение на случай временной зависимости состояний канала, приводящее к полумарковской модели. Для обоих случаев получены аналитические выражения для стационарных вероятностей состояний и сформулированы условия устойчивости системы, гарантирующие ограниченность очередей. Предлагается также модель для адаптивного управления пропускной способностью сети, оптимизирующая параметры передачи трафика в зависимости от текущего состояния канала, очереди запросов и скорости абонента. Эффективность предложенной модели подтверждается численным моделированием, демонстрирующим стабильную работу сети в широком диапазоне скоростей и параметров покрытия. Разработанная модель может применяться для оценки надёжности и оптимизации сетевых протоколов в условиях высокой мобильности абонентов, включая высокоскоростной железнодорожный транспорт, автомобильные сети и мобильные платформы.

Ключевые слова: мобильные сети, марковская модель, полумарковский процесс, устойчивость системы, преобразование Лапласа.

ВВЕДЕНИЕ

Современные системы связи, обеспечивающие устойчивую передачу данных от мобильных объектов, таких как высокоскоростные поезда, беспилотные транспортные средства, дроны и автомобили, активно используют инфраструктуру сотовых сетей общего пользования. Однако вне зон плотной застройки качество связи существенно снижается вследствие разреженности расположе-

ния базовых станций (БС), неполного покрытия и нестабильности радиоканала. Это проявляется в ослаблении сигнала, разрывах соединений и значительных задержках передачи данных.

При этом требования к качеству обслуживания (англ. *Quality of Service*, QoS) для абонентов в высокомобильной среде не уступают запросам пользователей в условиях стабильного покрытия: требуется устойчивая передача объёмного мультимедийного трафика, поддержка потокового видео и

передача данных в реальном времени с подвижных платформ. Такие сценарии особенно актуальны, например, при видеонаблюдении, телеметрии или автоматизированном управлении подвижными объектами.

Агрегация трафика является эффективным способом решения данных проблем [1, 2]. Существующие методы агрегации трафика (например, Link Aggregation Control Protocol (LACP), Multilink PPP) применяются в архитектурах LTE и LTE Advanced (LTE-A). В настоящей работе основное внимание уделяется сценариям, характерным для LTE и LTE-A, где агрегация реализуется на уровне протокола Packet Data Convergence Protocol (PDCP) и через механизмы Carrier Aggregation (CA). Модель может быть также адаптирована для сценариев 5G, однако специфика NR-архитектуры (New Radio) в данной работе не рассматривается подробно. Отметим, что данные методы демонстрируют высокую эффективность в сетях с устойчивым или умеренным покрытием, однако эффективность существенно падает при резких изменениях качества сигнала и высокой скорости движения абонентов, которая требует частого переключения между БС [2].

Анализ типовых сценариев функционирования сетей в условиях высокой мобильности абонентов позволяет выделить несколько фундаментальных проблем:

- Неполное покрытие сети. Вне городских территорий наблюдаются значительные мёртвые зоны, в которых уровень сигнала недостаточен для устойчивого соединения, что снижает общую пропускную способность сети и увеличивает потери пакетов.

- Динамика сигнала. Качество канала связи существенно зависит от траектории и скорости движения абонента. Доплеровский сдвиг, многолучевое распространение и неоднородность покрытия приводят к нестабильности канала и частым разрывам соединений.

- Неоптимальное поведение при потере связи. В случае разрыва соединения или ослабления сигнала инициируются многократные попытки повторной передачи пакетов, что приводит к перегрузке как канала, так и сетевой инфраструктуры, снижает общую эффективность передачи и увеличивает энергопотребление.

Существующие методы агрегации и управления мобильным трафиком недостаточно эффективно

справляются с перечисленными проблемами. Основные их недостатки таковы:

- Игнорирование временной корреляции. Большинство моделей предполагает независимые переходы между состояниями сети, что не учитывает инерционность оборудования и физические закономерности распространения сигнала.

- Отсутствие адаптации к топологии сети и скорости движения абонентов. В ряде подходов предполагается стационарная или квазистационарная среда, что делает их неприменимыми при высокой скорости движения и динамически изменяющемся качестве канала.

- Высокая латентность при переключении. Алгоритмы переключения между базовыми станциями могут приводить к длительным паузам в передаче трафика, особенно в условиях нестабильного сигнала.

Очевидным решением является улучшение существующих методов агрегации. Однако для оптимизации работы сети и поиска наилучшего решения требуется разработка соответствующих математических моделей. При этом существующие модели, как правило, либо являются упрощёнными, либо не учитывают ряд ключевых параметров реальных систем, таких как изменяющееся положение серверов и абонентов в пространстве [3, 4], временные корреляции в изменении состояния сети, динамическое перераспределение нагрузки, адаптивная буферизация данных; либо же учитывают пространственную геометрию элементов сети, но не учитывают сложную динамику очередей [5–7]; либо же анализируют исключительно проблемы связности компонентов сети, используя теорию графов [8] или методы теории перколяции [9–12], добавляя изменчивость состояний узлов во времени [13] и адаптируя классические методы статистической физики к задачам анализа беспроводных сетей [14, 15].

Цель данной работы – разработка новой модели управления трафиком в мобильных сетях, учитывающей стохастическую изменчивость пропускной способности каналов, временную корреляцию состояний сети и адаптивное распределение нагрузки для повышения устойчивости соединений и снижения задержек. В данной статье представлена стохастическая модель передачи агрегированного трафика, в которой характеристики канала описываются конечной марковской цепью, полученной путём дискретизации непрерывного сигнального



уровня [16]. Марковская модель оправдана ввиду экспоненциального характера статистики переходов между состояниями в реальных измерениях (например, при анализе сценариев переключения между базовыми станциями в LTE) [17].

В отличие от существующих моделей, предполагающих либо стационарность состояния канала, либо независимость переходов между состояниями, в данной работе предложена стохастическая модель, включающая временную корреляцию и пространственную структуру покрытия с учётом движения абонента между базовыми станциями по аффинной траектории. Основной вклад работы заключается:

- в построении марковской и полумарковской моделей канала связи на основе дискретизации уровня сигнала,
- в аналитическом выводе стационарных распределений состояний и условиях устойчивости очередей,
- в разработке адаптивной модели управления трафиком, оптимизирующей пропускную способность с учётом текущего состояния канала, скорости движения и длины очереди.

Работа имеет следующую структуру. В § 1 излагается процесс построения модели канала и представлена математическая модель адаптивного управления трафиком, выводятся основные уравнения модели, описывается марковский процесс переходов между состояниями и анализируются условия устойчивости системы. Формулировки и доказательства основных результатов даны в § 2. Здесь приводятся явные аналитические формулы для стационарного распределения, формулируется условие устойчивости модели. В § 3 представлены результаты численного моделирования предложенной модели в различных сценариях работы сети и осуществляется ее сравнение с существующими методами агрегации трафика. В заключении формулируются основные выводы по результатам исследования, оценивается эффективность предложенной модели и приводятся направления дальнейших исследований.

Настоящая работа ориентирована на архитектуры LTE и LTE-A, в которых агрегация трафика осуществляется на уровне PDCP и через механизмы Carrier Aggregation (CA). Рассматриваемая модель применима к сценариям с высокоскоростным перемещением (например, железнодорожного транспорта), характерным для данных технологий. Хотя в 5G NR применяются более сложные схемы

агрегации (включая Dual Connectivity и Service Data Adaptation Protocol (SDAP)), они выходят за рамки данной статьи и могут быть предметом последующего обобщения. Основное внимание сосредоточено на формализованном описании стохастической структуры радиоканала и адаптивного управления пропускной способностью в рамках LTE/LTE-A.

1. ПОСТРОЕНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Мобильную сеть будем рассматривать как совокупность базовых станций и мобильных абонентов, перемещающихся по маршрутам с различным уровнем покрытия сигнала. При этом под мобильным абонентом будем понимать единый агрегатор или сервер, формирующий объединённый трафик от множества устройств на борту высокоскоростного поезда или автомобиля. Практическая реализация такой схемы получила распространение в современных сетях 4G/5G (примером могут служить мультирелейные системы, в рамках которых поезд или автобус имеет единую точку доступа к сотовой сети) [1]. Это позволяет агрегировать трафик множества устройств в один канал и упростить процесс переключения между базовыми станциями, в то время как отдельные устройства внутри вагона подключаются к этому «серверу» [1, 18, 19]. В контексте LTE- и LTE-A-архитектур агрегатор реализуется в виде абонентского терминала или мультирелейной платформы, объединяющей трафик с борта транспортного средства и взаимодействующей с несколькими компонентными сетями (CA).

Будем предполагать, что покрытие сети неплотное, т. е. БС располагаются на таком удалении друг от друга, чтобы имели место разрывы в покрытии, а скорость движения абонентов достаточно высокая, чтобы для передачи необходимого объема трафика нужно было задействовать последовательно несколько БС.

1.1. Предположения модели

Топологию сети естественно моделировать в виде ориентированного графа $G = (V, E)$ [3, 5, 8], где V – множество узлов сети (например, базовых станций); $E \subseteq V \times V$ – множество ориентированных рёбер, представляющих возможные маршруты передачи данных между узлами.

Сделаем несколько предположений относительно рассматриваемой модели, которые существенно упрощают анализ, но при этом не ограничивают общность подхода и основные качественные выводы.

• **Граф базовых станций является линейным:** будем предполагать, что БС расположены вдоль прямой линии в точках с координатами

$$x_j = j \Delta x, \quad j \in \mathbb{Z}_+,$$

где Δx – фиксированное расстояние между соседними БС. Другими словами, предполагается, что БС расположены равномерно в узлах одномерной решётки $\mathbb{Z}_+$. Такое предположение отражает типичную ситуацию покрытия вдоль транспортных магистралей или железных дорог.

Замечание. Следует подчеркнуть, что рассматриваемая линейная расстановка БС с равномерным шагом вдоль прямой является идеализированным допущением, характерным в первую очередь для магистральных сценариев (железнодорожных, автомагистралей и т. д.). В реальных системах могут встречаться вариации во взаимном расположении БС, участки трассы с кривизной, зоны неоднородного рельефа (туннели, мосты и пр.). Тем не менее, линейная модель даёт удобный базовый каркас для анализа основных закономерностей, особенно при движении по относительно прямым участкам пути. Это упрощение не влияет на общую логику методов и выводы, но при необходимости может быть уточнено.

• **Движение агрегатора равномерно:** агрегатор перемещается вдоль указанной прямой линии с постоянной скоростью:

$$v(t) = v_0, \quad v_0 > 0.$$

Это предположение оправдано при моделировании стандартных сценариев (например, движения высокоскоростных поездов или автомобилей на автомагистралях), в которых изменения скорости относительно малы и агрегатор движется стабильно.

Замечание. В работе предполагается постоянная скорость агрегатора $v(t) = v_0$ (или нормированная единичная скорость). Такой выбор оправдан стационарностью условий высокоскоростного транспорта на длинных магистралях. При этом реальные сценарии могут включать участки разгона (торможения), а также колебания скорости. Если такие эффекты существенны, модель можно расширить. В данном же исследовании сосредоточимся на базовом сценарии с постоянной или почти постоянной скоростью, что обеспечивает перио-

дичность движения и упрощает дальнейшие аналитические выкладки без потери общности.

• **Влияние БС локализовано:** качество соединения в каждой точке траектории агрегатора определяется только двумя ближайшими БС – текущей (позади агрегатора) и следующей (перед агрегатором). Такое упрощение обосновано быстротой затухания сигнала с расстоянием и позволяет пренебречь влиянием дальних БС без существенной потери точности.

Таким образом, при условии равномерного расположения базовых станций и равномерного движения агрегатора процесс перемещения агрегатора можно моделировать как периодическое движение точки по полуинтервалу $[0, 1)$. Можно считать также, что при достижении агрегатором положения 1 происходит смена текущей и следующей базовых станций, и агрегатор снова начинает движение из точки 0. Такая интерпретация существенно упрощает математический анализ и численное моделирование динамики сети.

С учётом этих предположений состояние агрегатора в любой момент времени может быть полностью параметризовано его положением на аффинном полуинтервале $[0, 1)$, а исследование модели можно ограничить периодом прохождения данного расстояния.

Таким образом, при условии равномерного расположения базовых станций и равномерного движения агрегатора его движение можно строго описать следующим образом.

Обозначим через $T = \frac{\Delta x}{v_0}$ время прохождения

агрегатором расстояния Δx между двумя соседними БС.

Тогда положение агрегатора в любой момент времени t моделируется с помощью периодического аффинного параметра $s(t) \in [0, 1)$, задаваемого формулой

$$s(t) = \frac{t \bmod T}{T},$$

где $s(t) = 0$ соответствует положению агрегатора ровно в точке текущей базовой станции, а $s(t) = 1$ – достижению следующей БС.

При достижении агрегатором положения $s(t) = 1$ он мгновенно начинает движение на новом отрезке, где следующая базовая станция становится текущей, а новая соседняя базовая станция – следующей.

Таким образом, движение агрегатора по сети представляет собой периодический процесс, кото-



рый в рамках модели можно представить как повторяющееся движение точки по полуинтервалу $[0, 1]$:

$$x(t) = s(t).$$

Важным частным случаем, который позволяет максимально упростить модель и последующий анализ, является выбор единичной скорости агрегатора $v_0 = 1$ и единичного расстояния между базовыми станциями $\Delta x = 1$.

В этом случае время прохождения между соседними БС становится единичным ($T = 1$), и положение агрегатора $x(t)$ на полуинтервале $[0, 1]$ задаётся простой периодической зависимостью

$$x(t) = t \bmod 1.$$

Таким образом, агрегатор перемещается по полуинтервалу $[0, 1]$ с постоянной единичной скоростью и, достигая точки 1, мгновенно переходит к следующему полуинтервалу, начиная вновь с позиции 0. Графически эта функция представляет собой периодическую пилообразную зависимость с единичным периодом.

Таким образом, положение агрегатора на полуинтервале $[0, 1]$ полностью определяет его удалённость от двух соседних базовых станций.

Ввиду периодичности модели для анализа свойств системы достаточно рассмотреть её поведение на одном фиксированном периоде $t \in [0, 1]$. Результаты на остальных временных интервалах будут аналогичны вследствие указанной периодичности.

1.2. Динамика расстояний до базовых станций и качества сигнала с учётом топологии сети

Расстояния от агрегатора до текущей (позади него) и следующей (впереди него) базовых станций при положении $x(t) \in [0, 1]$ задаются следующими формулами:

$$d_{\text{current}}(t) = x(t), d_{\text{next}}(t) = 1 - x(t), 0 \leq x(t) \leq 0,5;$$

$$d_{\text{current}}(t) = 1 - x(t), d_{\text{next}}(t) = x(t), 0,5 < x(t) \leq 1.$$

Из этих формул явно следует, что:

- в начале периода ($x(t) = 0$) агрегатор находится непосредственно на текущей базовой станции (расстояние до неё равно нулю), а до следующей станции расстояние максимально (равно единице);
- в середине периода ($x(t) = 0,5$) агрегатор находится ровно посередине между двумя соседними базовыми станциями, и расстояния до них равны 0,5;

- в конце периода ($x(t) = 1$) агрегатор достигает следующей базовой станции, после чего происходит мгновенный переход на новый отрезок, где следующая базовая станция становится текущей, и положение агрегатора снова становится равным нулю.

Таким образом, за один период агрегатор сначала удаляется от текущей базовой станции (при движении от 0 до 0,5), а затем приближается к следующей станции (при движении от 0,5 до 1). В результате расстояние до ближайшей базовой станции за один период сначала монотонно возрастает от 0 до 0,5, а затем монотонно убывает обратно до нуля.

Дополнительно предположим, что каждая базовая станция характеризуется конечным радиусом действия R , в пределах которого агрегатор получает устойчивый сигнал (состояние S_1). Кроме того, за пределами радиуса R существует небольшая переходная зона шириной δ , в которой сигнал ослаблен (состояние S_2).

Таким образом, для каждой базовой станции выделяются три зоны покрытия:

- зона устойчивого сигнала (состояние S_1): $0 \leq d \leq R$,
- переходная зона ослабленного сигнала (состояние S_2): $R < d \leq R + \delta$,
- мёртвая зона (состояние S_3): $d > R + \delta$, где d – текущее расстояние до базовой станции

$$d = d(t) = \min \{d_{\text{current}}(t), d_{\text{next}}(t)\}.$$

Предполагается, что выполнено следующее естественное условие:

$$2(R + \delta) < 1,$$

т. е. суммарный диаметр зон покрытия одной базовой станции (зона устойчивого сигнала плюс переходная зона с обеих сторон) существенно меньше расстояния между соседними базовыми станциями. Это гарантирует наличие мёртвых зон между зонами покрытия соседних БС, где сигнал полностью отсутствует.

При движении агрегатора по полуинтервалу $[0, 1]$ агрегатор последовательно проходит:

- через зону устойчивого сигнала текущей станции,
- через переходную зону ослабленного сигнала текущей станции,
- через мёртвую зону,
- через переходную зону ослабленного сигнала следующей станции,
- через зону устойчивого сигнала следующей станции.

В общем случае влияние скорости движения и расстояния до ближайшей базовой станции на качество соединения моделируется специальной функцией потерь, объединяющей линейные и нелинейные эффекты [20]:

$$L(d(t), v(t)) = L_0 + ad(t)^\gamma + k_1v(t) + k_2\log(1+v(t)) + k_3v(t)^\alpha,$$

где L_0 – базовые потери при идеальном случае (нулевое расстояние и нулевая скорость); $ad(t)^\gamma$ – компонент затухания сигнала, зависящий от расстояния $d(t)$ до базовой станции (степенная модель распространения сигнала, обычно $\gamma \in [2, 4]$; $k_1v(t)$ – линейный компонент, связанный с доплеровским сдвигом частоты; $k_2\log(1+v(t))$ – логарифмическое затухание, вызванное многолучевыми замираниями (*fading*); $k_3v(t)^\alpha$ – нелинейные эффекты, возникающие при переключении между базовыми станциями и высоких скоростях ($\alpha > 1$).

Таким образом, представленная модель одновременно учитывает влияние обоих ключевых факторов – удалённости от базовой станции и скорости движения агрегатора – что делает её физически адекватной и практически полезной для дальнейшего анализа.

Линейная модель $L(v) = L_0 + ad(t)^\gamma + k_1v(t)$ сохраняется как частный случай ($k_2 = k_3 = 0$), обеспечивая аналитическую простоту для базового анализа [20], и в рамках данной работы будет рассматриваться именно данная модель.

1.3. Процесс переходов между дискретными уровнями сигнала в общем случае

В рамках предложенной модели агрегатор последовательно проходит три зоны с различным качеством сигнала (устойчивый сигнал, ослабленный сигнал и мёртвая зона). В реальных сетях переходы между состояниями соединения (S_1, S_2, S_3) происходят не мгновенно, а характеризуются некоторой задержкой и инерционностью. Основными причинами таких эффектов являются [20]:

- Постоянство внешних условий: параметры канала связи остаются относительно стабильными на протяжении некоторого времени, поэтому переходы между состояниями происходят постепенно, а не резко.

- Инерционность оборудования: переключение между базовыми станциями требует дополнительного времени на согласование и обработку сигналов.

- Физическая траектория движения: при движении по определённой траектории переходы между состояниями связаны с топологией сети и регулярностью расположения базовых станций.

Эти эффекты создают временную корреляцию между состояниями канала, что означает зависимость вероятностей переходов не только от текущего состояния, но и от времени пребывания в нём. Строгое математическое описание таких ситуаций приводит к необходимости рассмотрения более общего класса процессов, называемых полумарковскими (англ. *Semi-Markov Processes*, SMP) [21, 22].

Рассмотрим случайный процесс $X(t)$, моделирующий изменение уровня сигнала в мобильной сети. Этот процесс представляет собой полумарковскую модель с тремя состояниями: S_1 (устойчивый сигнал), S_2 (ослабленный сигнал), S_3 (мёртвая зона). Моменты переходов образуют возрастающую последовательность $\{\tau_n\}$ случайных величин, при этом $X(t)$ сохраняет текущее состояние между переходами [23].

Полумарковская модель уровней сигнала задаётся следующей системой интегро-дифференциальных уравнений для вероятностей состояний $p_i(t) = P\{X(t) = S_i\}$, $i = 1, 2, 3$:

$$\begin{aligned} \frac{dp_i(t)}{dt} = & \sum_{j \neq i} \int_0^t p_j(\tau) \alpha_{ji}(t-\tau) d\tau - \\ & - \sum_{k \neq i} \int_0^t p_i(\tau) \alpha_{ik}(t-\tau) d\tau, \end{aligned} \quad (1)$$

где интенсивность перехода задаётся функцией

$$\begin{aligned} \alpha_{ij}(t) = & \lambda_{ij}(t) \exp\left(-\int_0^t \lambda_i(u) du\right), \\ \lambda_i(t) = & \sum_{k \neq i} \lambda_{ik}(t), \quad i = 1, 2, 3, \end{aligned}$$

Будем полагать, что заданы фиксированные начальные условия $p_1(0) = 1, p_2(0) = p_3(0) = 0$.

Для решения системы (1) удобно применять преобразование Лапласа [24]. При этом в пространстве Лапласа получаем систему алгебраических уравнений



$$\begin{aligned} \tilde{s}p_i(s) - p_i(0) = \\ = \sum_{j \neq i} \tilde{p}_j(s) \tilde{\alpha}_{ji}(s) - \sum_{k \neq i} \tilde{p}_i(s) \tilde{\alpha}_{ik}(s). \end{aligned}$$

В общем случае времена пребывания могут иметь произвольные распределения $F_i(t)$: $T_i \sim F_i(t)$. Тогда система уравнений (1) становится достаточно сложной для аналитического решения. В такой ситуации можно применять численные схемы по временной сетке $\{t_n\}$, например:

$$\begin{aligned} p_i(t_{n+1}) = p_i(t_n) + \Delta t \left[\sum_{j \neq i} \sum_{m=0}^n p_j(t_m) \alpha_{ji}(t_{n+1} - t_m) - \right. \\ \left. - \sum_{k \neq i} \sum_{m=0}^n p_i(t_m) \alpha_{ik}(t_{n+1} - t_m) \right]. \end{aligned}$$

При более сложных распределениях (гамма-, Вейбулла и т. п.) такой прямой численный подход позволяет эффективно получать решения.

На практике часто используются следующие распределения времён пребывания в состояниях: гамма- (Эрланга), Вейбулла и равномерное распределения. Ниже приведены распределения времён пребывания и соответствующие выражения для интенсивностей переходов $\alpha_{ij}(t) = \lambda_{ij}(t) \times \exp\left(-\int_0^t \lambda_i(u) du\right)$, а также их преобразования Лапласа для данных распределений $\tilde{\alpha}_{ij}(s) = \mathcal{L}\{\alpha_{ij}(t)\}$:

• Гамма-распределение (распределение Эрланга):

Функция интенсивности, интенсивность перехода в соседнее состояние и преобразование Лапласа соответственно имеют следующий вид:

$$\lambda_i(t) = \frac{\lambda_i^k t^{k-1} e^{-\lambda_i t}}{(k-1)! - \sum_{m=0}^{k-1} \frac{(\lambda_i t)^m}{m!}},$$

$$\alpha_{ij}(t) = p_{ij} \lambda_i(t), \quad \tilde{\alpha}_{ij}(s) = p_{ij} \frac{\lambda_i^k}{(s + \lambda_i)^k}.$$

• Распределение Вейбулла:

$$\lambda_i(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1}, \quad \alpha_{ij}(t) = p_{ij} \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right],$$

$$\tilde{\alpha}_{ij}(s) = p_{ij} \frac{\beta}{\alpha^\beta} \int_0^\infty t^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\beta - st\right] dt.$$

• Равномерное распределение на отрезке $[a, b]$:

$$\lambda_i(t) = \begin{cases} 0, & t < a, \\ \frac{1}{b-t}, & a \leq t < b, \\ \infty, & t = b, \end{cases}$$

$$\alpha_{ij}(t) = \begin{cases} 0, & t < a, \\ \frac{p_{ij}}{b-a}, & a \leq t \leq b, \\ 0, & t > b, \end{cases}$$

$$\tilde{\alpha}_{ij}(s) = p_{ij} \frac{e^{-as} - e^{-bs}}{(b-a)s}.$$

Подстановка вышеописанных параметров в систему уравнений (1) позволяет получить явные выражения для вероятностных характеристик модели при различных распределениях времени пребывания в состояниях, а также вывести важнейшие метрики качества функционирования системы.

Однако аналитическое исследование таких моделей в общем случае приводит к громоздким выражениям, затрудняющим их интерпретацию и применение. Поэтому в дальнейшем ограничимся рассмотрением марковского случая, предполагающего экспоненциальное распределение времени пребывания в каждом состоянии:

$$\mathbb{P}(T_i > t) = e^{-\lambda_i t},$$

что соответствует отсутствию памяти и существенно упрощает анализ. Такое допущение означает, что временные корреляции и инерционность каналов считаются усреднёнными и аппроксимируются постоянными интенсивностями переходов между состояниями.

Это упрощение позволяет получить строго аналитические и легко интерпретируемые результаты, которые и будут изложены в § 2 как основные аналитические следствия модели.

2. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

2.1. Механизм оптимального управления

Для адаптивной корректировки параметров передачи трафика (формально) введём функцию $u(t)$, зависящую от текущего состояния сети:

$$u(t) = \begin{cases} \mu_{\text{norm}}, & \text{если } X(t) = S_1, \\ \mu_{\text{slow}} g(\tau_2, \theta), & \text{если } X(t) = S_2, \\ 0, & \text{если } X(t) = S_3, \end{cases}$$

где $g(\tau_2, \theta)$ – адаптивная функция, зависящая от времени пребывания в состоянии S_2 и параметров θ , подлежащих оптимизации.

Для оптимизации работы системы формулируется целевая функция

$$J(u) = \int_0^T [\omega_1 L_{\text{pack}}(u(t)) + \omega_2 D(u(t))] dt, \quad (2)$$

где $L_{\text{pack}}(u(t))$ – функция, отражающая потери пакетов в зависимости от выбранной стратегии $u(t)$; $D(u(t))$ – функция задержек передачи данных; ω_1 и ω_2 – коэффициенты важности соответствующих критериев; T – рассматриваемый временной интервал.

Оптимальная стратегия $u^*(t)$ достигается путём минимизации целевой функции (2). В качестве метода оптимизации можно использовать, например, градиентный спуск или методы динамического программирования [25, 26]. Для градиентного спуска по параметрам θ корректировка осуществляется по хорошо известному алгоритму [25, 27]

$$\theta_{n+1} = \theta_n - \eta \nabla_{\theta} J(u),$$

где η – параметр алгоритма, а $\nabla_{\theta} J(u)$ – градиент целевой функции по параметрам θ . Подобные подходы широко применяются в задачах калибровки систем связи и управления потоками в сетях.

Кроме управления скоростью передачи, функция $u(t)$ может влиять и на переходные интенсивности между состояниями сети. Например, можно модифицировать интенсивности так, чтобы уменьшить вероятность разрыва связи или ускорить восстановление:

$$\lambda_{ij}(v, u(t)) = \lambda_{ij}^0 h(u(t)) e^{-k_v v(t)}, \quad (3)$$

где λ_{ij}^0 – базовые интенсивности; функция $h(u(t))$ описывает корректирующий эффект управления, а $e^{-k_v v(t)}$ отражает влияние скорости движения $v(t)$ агрегатора (при необходимости). Подобные поправки позволяют моделировать физические и протокольные механизмы, при которых, например, более агрессивное использование ресурса (большая мощность передачи) увеличивает значение $h(u(t))$ и тем самым снижает вероятность перехода в мёртвую зону.

В рамках данной работы используется экспоненциальная форма управляющей функции (3) как

базовая. Она хорошо отражает интуитивно ожидаемое экспоненциальное ухудшение параметров канала при длительном нахождении в ослабленном состоянии [28], выбор экспоненты оправдан аналитической простотой, применимостью в известных моделях каналов с затуханием возможностью калибровки параметра α в рамках процедуры оптимизации [29].

Однако возможны и другие виды функций (логистическая, линейная и др.), в зависимости от физических или протокольных ограничений. Подобный выбор функции $g(\tau_2, \theta)$ может быть обоснован с помощью данных или в рамках аппроксимации механизмов, реализуемых, например, в адаптивной модуляции современных протоколов связи (см. LTE, 5G [30]).

В настоящей работе целевая функция (2) рассматривается в общем виде, однако для численного моделирования можно конкретизировать её следующим образом:

$$L_{\text{pack}} = \frac{1}{T} \int_0^T \mathbb{I}_{\text{loss}}(t) dt,$$

где $L_{\text{pack}}(u(t))$ – доля потерянных пакетов, $\mathbb{I}_{\text{loss}}(t) = 1$, если пакет в момент t был отброшен, и $\mathbb{I}_{\text{loss}}(t) = 0$ в противном случае.

Усреднённая задержка передачи пакетов $D(u(t))$ может быть оценена, например, по следующей формуле:

$$D(u(t)) = \frac{1}{N} \sum_{i=1}^N W_i,$$

где W_i – суммарное время пребывания i -го пакета в системе.

Таким образом, введение управления $u(t)$ и целевой функции даёт возможность комплексного подхода к оптимизации в сетях мобильной связи [31]: как регулирования скорости передачи (или других параметров канала), так и управления вероятностью разрывов и восстановлений путём изменения интенсивностей λ_{ij} . На практике выбор конкретной формы функции $u(t)$ и метода оптимизации определяется требованиями к качеству (снижению задержек, потерь) и доступными аппаратными (протокольными) возможностями [32].

Кроме того, подобное управление может быть реализовано посредством дополнительного модуля (платы) с программным управлением, в котором, например, применяются нейронные сети для прогнозирования состояния канала и динамической



настройки параметров. Тем самым удаётся вовремя реагировать на ухудшения связи и заранее адаптировать скорость передачи и вероятность разрывов в реальном времени.

В зависимости от специфических требований к обслуживанию (например, ограничения на задержку, порог вероятности потерь или ограничения на энергоёмкость передачи) предложенная целевая функция (2) может быть уточнена путём добавления дополнительных критериев или штрафов за превышение норм качества. Таким образом, при решении конкретных инженерных задач форма $L_{\text{pack}}(\cdot)$ и $D(\cdot)$ может быть выбрана с учётом практической значимости конкретных метрик качества работы сети (пакетные потери, время простоя и пр.). В настоящей работе функция $J(u)$ остаётся в общем виде; в разделе, посвящённом частным предположениям и упрощениям, продемонстрируем, как именно происходит выбор целевой функции для специальных случаев.

2.2. Основные свойства системы

Теперь рассмотрим основные характеристики системы. Для простоты и прозрачности дальнейшего анализа предположим, что интенсивности переходов между состояниями связи постоянны в пределах каждой зоны и зависят только от текущего состояния, в котором находится агрегатор. Это означает, что переходы между уровнями качества сигнала можно описывать с помощью фиксированных параметров, отражающих физические характеристики среды и динамику перемещения.

Таким образом, модель использует конечное число постоянных интенсивностей переходов, каждая из которых характеризует типичный режим перехода между зонами покрытия базовой станции. При предположении об экспоненциальном распределении времени пребывания в каждом состоянии модель сводится к непрерывному марковскому процессу с конечным числом состояний, задаваемому системой уравнений Колмогорова и, таким образом, вместо рассмотрения системы (1) можно перейти к рассмотрению инфинитезимальной матрицы переходов

$$Q = \begin{pmatrix} -\lambda_{12} & \lambda_{12} & 0 \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23}) & \lambda_{23} \\ 0 & \lambda_{32} & -\lambda_{32} \end{pmatrix},$$

а вектор вероятностей состояний $p(t) = (p_1(t), p_2(t), p_3(t))$ удовлетворяет системе уравнений Колмогорова

$$\frac{dp(t)}{dt} = p(t)Q.$$

Здесь коэффициенты λ_{ij} имеют следующую физическую интерпретацию:

- λ_{12} – интенсивность перехода от устойчивого сигнала S_1 к ослабленному S_2 , соответствующая движению из зоны $d \leq R$ в зону $R < d \leq R + \delta$;
- λ_{21} – интенсивность восстановления сигнала от состояния S_2 к S_1 при обратном переходе в зону хорошего покрытия;
- λ_{23} – интенсивность потери сигнала (переход от состояния S_2 к S_3), что соответствует попаданию в мёртвую зону;
- λ_{32} – интенсивность восстановления сигнала при возвращении из мёртвой зоны S_3 в переходную зону S_2 .

Замечание. В описываемом подходе структура сети явно не входит в модель. Однако она может быть косвенно учтена путём параметризации матрицы Q на основе матрицы смежности графа покрытия или других структурных характеристик сети. Такой подход открывает возможность для обобщения модели на случай пространственно неоднородных сетей или реальных телекоммуникационных инфраструктур.

Теорема 1 (о существовании кусочно-гладкой управляющей функции). Для канала, находящегося в одном из трёх состояний: S_1 (устойчивое состояние), S_2 (ослабленное состояние) и S_3 (разрыв связи), существует кусочно-гладкая управляющая функция общего вида $u(t)$, зависящая только от текущего состояния канала $X(t)$ такая, что

$$u(t) = \begin{cases} u_1, & X(t) = S_1, \\ u_2, & X(t) = S_2, \\ u_3, & X(t) = S_3, \end{cases}$$

где u_i – параметры управления в состоянии S_i ($i=1, 2, 3$). Интенсивности обслуживания $\mu_i(u(t))$ в каждом состоянии канала и переходов $\lambda_{ij}(u(t))$ между состояниями при этом корректно определены и управляемы на каждом интервале постоянства состояния.

Данная функция управления определяет интенсивность обслуживания $\mu_i(u(t))$ в каждом состоянии канала, а также может влиять на интенсивно-

сти переходов $\lambda_{ij}(u(t))$ между состояниями (например, путём управляемого ускорения восстановления канала или замедления его ухудшения). Иными словами, когда канал находится в состоянии S_i , система применяет управление u_i , которое задаёт текущую скорость обслуживания $\mu_i(u_i)$ и может изменять переходные интенсивности $\lambda_{ij}(u_i)$ в другие состояния.

Д о к а з а т е л ь с т в о (обоснование выбора формы $u(t)$). Управление, зависящее только от текущего состояния $X(t)$, естественно в контексте адаптивных систем связи. Такая кусочно-заданная функция $u(t)$ позволяет мгновенно реагировать на изменения качества канала: при переходе из устойчивого состояния в ослабленное или при разрыве связи управление скачкообразно меняет свое значение на новое, соответствующее ухудшившимся условиям. Это обеспечивает адаптивное регулирование обслуживания – например, при ухудшении канала (S_2) можно уменьшить скорость обслуживания или включить более надёжный режим передачи, а при разрыве (S_3) – инициировать процедуры восстановления связи (что эквивалентно увеличению интенсивности перехода обратно в состояние S_1 или S_2). В устойчивом состоянии S_1 управление может возвращаться в режим максимальной пропускной способности. Такая форма функции $u(t)$ гарантирует, что система динамически подстраивается под текущее состояние канала, улучшая устойчивость обслуживания при деградации канала и сокращая длительность простоев при разрывах. ♦

Теорема 2 (критерий устойчивости системы с очередью). Рассмотрим марковский процесс $X(t)$, принимающий три состояния $\{S_1, S_2, S_3\}$, где состояния S_1 и S_2 имеют интенсивности обслуживания μ_1 и μ_2 соответственно, а в состоянии S_3 передача невозможна ($\mu_3 = 0$). Пусть λ_{in} – интенсивность входного (пуассоновского) потока заявок. Обозначим через π_1, π_2, π_3 стационарные вероятности пребывания канала в состояниях S_1, S_2, S_3 (если они существуют).

Система очередей (с неограниченным буфером) и переменным качеством сигнала устойчива тогда и только тогда, когда

$$\lambda_{in} < \pi_1 \mu_1 + \pi_2 \mu_2. \quad (4)$$

Условие (4) означает, что средняя входная нагрузка должна быть меньше средней пропускной способности канала, равной $\pi_1 \mu_1 + \pi_2 \mu_2$. При

нарушении условия (4) очередь (при неограниченной ёмкости) возрастает неограниченно, см. работы [3, 24].

Д о к а з а т е л ь с т в о. В модели $M/M/1$ условие устойчивости $\lambda_{in} < \mu$ есть необходимое и достаточное [2]. Для случая с переменной μ , зависящей от состояния канала, средняя скорость обслуживания равна $\sum_{i=1}^3 \pi_i \mu_i$. Так как $\mu_3 = 0$, вклад состояния S_3 в пропуск-

ную способность нулевой. Следовательно, критерий переходит в неравенство (4). При выполнении условия (4) существует единственное стационарное распределение размера очереди; иначе стационарное распределение отсутствует.

Стационарные вероятности π_1, π_2, π_3 определяются из уравнений Колмогорова (1). Стационарный вектор $\pi = (\pi_1, \pi_2, \pi_3)$ удовлетворяет системе

$$\begin{cases} -\lambda_{12}\pi_1 + \lambda_{21}\pi_2 = 0, \\ \lambda_{12}\pi_1 - (\lambda_{21} + \lambda_{23})\pi_2 + \lambda_{32}\pi_3 = 0, \\ \pi_1 + \pi_2 + \pi_3 = 1. \end{cases}$$

Решая систему, получаем:

$$\pi_2 = \frac{\lambda_{12}}{\lambda_{21}} \pi_1, \quad \pi_3 = \frac{\lambda_{23}}{\lambda_{32}} \pi_2 = \frac{\lambda_{12}\lambda_{23}}{\lambda_{21}\lambda_{32}} \pi_1.$$

Подставляя в условие нормировки:

$$\begin{aligned} \pi_1 + \pi_2 + \pi_3 &= \pi_1 \left(1 + \frac{\lambda_{12}}{\lambda_{21}} + \frac{\lambda_{12}\lambda_{23}}{\lambda_{21}\lambda_{32}} \right) = 1 \Rightarrow \\ \Rightarrow \pi_1 &= \left(1 + \frac{\lambda_{12}}{\lambda_{21}} + \frac{\lambda_{12}\lambda_{23}}{\lambda_{21}\lambda_{32}} \right)^{-1}. \end{aligned}$$

Подстановка π_1, π_2, π_3 в неравенство (4) даёт явное численное условие

$$\lambda_{in} \leq \left(\frac{\lambda_{21}\lambda_{32}}{\Delta} \right) \mu_1 + \left(\frac{\lambda_{12}\lambda_{32}}{\Delta} \right) \mu_2,$$

где

$$\Delta = \lambda_{21}\lambda_{32} + \lambda_{12}\lambda_{32} + \lambda_{12}\lambda_{23}. \quad \blacklozenge$$

Следствие (условие устойчивости системы с управлением). В стационарном режиме условие устойчивости системы с учётом оптимального управления записывается как

$$\lambda_{in} < \pi_1(u^*) \mu_{\text{norm}} + \pi_2(u^*) \mu_{\text{slow}},$$

где стационарные вероятности $\pi_i(u^*)$ зависят от оптимальной стратегии $u^*(t)$.



2.3. Операционные характеристики системы в стационарном режиме

Хотя в предыдущих разделах изложение было сосредоточено на описании канала (состояний S_1, S_2, S_3) и его переходных интенсивностях, в реальной системе с очередью полное состояние в момент времени t задаётся парой «состояние канала – длина очереди».

Умея оценивать эволюцию канала в терминах состояний $\{S_1, S_2, S_3\}$ и управляющей функции $u(t)$, можно оценить качество работы системы с учётом трафика в реальных системах. Для этого объединим марковскую модель канала с моделью очереди типа $M/M/1$ (либо $M/M/1/N$ при конечном буфере), где скорость обслуживания зависит от текущего состояния канала и, при необходимости, от управления $u(t)$.

Такой подход основан на классических результатах теории массового обслуживания [3, 4], применённых к ситуации, когда параметры обслуживания (и даже переходы канала из одного состояния в другое) могут меняться в зависимости от стратегии управления. Как как будет показано далее, в стационарном режиме ключевые вероятностные характеристики – вероятность потерь, среднее число заявок в системе, пропускная способность – выражаются через стандартные формулы Эрланга и Литтла, но с учётом средних (эффективных) скоростей, задаваемых состояниями канала.

Рассмотрим систему с очередью, которую обслуживает управляемый марковский канал с тремя состояниями $\{S_1, S_2, S_3\}$, где управление $u(t)$ может влиять как на интенсивность обслуживания $\mu_i(u)$ в состояниях S_i , так и на интенсивности переходов $\lambda_{ij}(u)$. Предположим следующее:

- Очередь имеет конечную ёмкость: максимум N пакетов в ожидании (значит, всего $N+1$ пакетов вместе с обслуживаемым).

- В стационарном режиме канал проводит долю времени $\pi_i(u^*)$ в состоянии S_i (при оптимальном или фиксированном управлении u^*).

- Если $\mu_3(u) = 0$ (пребывание в «мёртвой» зоне), то $\bar{\mu} = \pi_1(u^*)\mu_1(u^*) + \pi_2(u^*)\mu_2(u^*)$ есть эффективная (средняя) скорость обслуживания.

- Интенсивность входного потока заявок – λ_{in} .

Тогда с учётом предположений можно найти операционные характеристики работы системы.

- *Коэффициент загрузки.* Определим относительную нагрузку (загрузку):

$$\rho = \frac{\lambda_{in}}{\bar{\mu}} = \frac{\lambda_{in}}{\pi_1(u^*)\mu_1(u^*) + \pi_2(u^*)\mu_2(u^*)}.$$

Если $\rho < 1$ при данном u^* , система может работать без «коллапса» в неограниченном буфере; если $\rho \geq 1$, очередь растёт или происходят большие потери.

- *Вероятность потерь.* Для конечного N вероятность потери пакета при переполнении очереди задаётся (аналогично формуле Эрланга) через

$$P_{loss} = P\{\text{очередь} = N\} = \frac{\rho^{N+1}}{\sum_{k=0}^{N+1} \rho^k},$$

где ρ определяется через $\pi_i(u^*)$ и $\mu_i(u^*)$. При $\rho < 1$ и $N \rightarrow \infty$ получаем $P_{loss} \rightarrow 0$, а при $\rho > 1$ имеем $P_{loss} \rightarrow 1$.

- *Среднее число пакетов и задержка.* Обозначим $L = E[n]$ математическое ожидание числа пакетов в очереди и на обслуживании. При $\rho \neq 1$ справедлива формула

$$L = \frac{\rho}{1-\rho} - \frac{(N+2)\rho^{N+2}(1-\rho)}{1-\rho^{N+2}}.$$

Среднее число пакетов в очереди L_q получается вычитанием средней занятой позиции обслуживания; аналогично для среднего времени пребывания (W) и ожидания (W_q) применяем закон Литтла:

$$L = \lambda_{eff} W, \quad L_q = \lambda_{eff} W_q,$$

где $\lambda_{eff} = \lambda_{in}(1 - P_{loss})$ – интенсивность фактического потока поступающих и обслуживаемых пакетов.

- *Средняя пропускная способность.* В стационарном режиме система пропускает поток с интенсивностью

$$\lambda_{out} = \lambda_{eff} = \lambda_{in}(1 - P_{loss}).$$

Если $\rho < 1$ и $N \rightarrow \infty$, то $\lambda_{out} = \lambda_{in}$ (нет потерь). При $\rho > 1$ очередь переполнена и $\lambda_{out} \approx \bar{\mu}$.

2.4. Следствия и замечания

- *Граничные случаи (сводимость к классическим моделям).* Если канал практически всегда

находится в хорошем состоянии S_1 (т. е. $\pi_1 \approx 1$, $\pi_2 \approx \pi_3 \approx 0$, $\mu_1 = \mu$), то вся система сводится к классическому $M/M/1$. Тогда критерий устойчивости принимает вид $\lambda_{in} < \mu$, а формулы для вероятности потерь, среднего числа пакетов и времени ожидания упрощаются до хорошо известных результатов (формула Эрланга, формулы Литтла и т. д.). Напротив, если значительную долю времени агрегатор пребывает в мёртвой зоне S_3 (π_3 велика, $\mu_3 \approx 0$), то эффективная скорость $\bar{\mu}$ резко падает, и система ведёт себя как «медленный сервер» с высокой нагрузкой ρ .

• **Пример влияния пребывания в мёртвой зоне на потери.** Пусть $\mu_1 = 1$, $\mu_2 = 0,5$, $\mu_3 = 0$, а доли времени в состояниях таковы: $\pi_1 = 0,8$, $\pi_2 = 0,15$, $\pi_3 = 0,05$. Тогда $\bar{\mu} = 0,8 \cdot 1 + 0,15 \cdot 0,5 = 0,875$. При входном потоке $\lambda_{in} = 0,7$ имеем $\rho \approx 0,8 < 1$, и система работает без существенных потерь. Однако при $\lambda_{in} = 1,0$ ($\rho \approx 1,14 > 1$) очередь начинает неограниченно расти (при большом буфере) либо достигает максимума N (при конечном буфере); возникает ощутимая вероятность потерь ($P_{loss} \approx 0,2$), а выходная пропускная способность насыщается на уровне $\lambda_{out} \approx \bar{\mu} = 0,875$.

• **Физический смысл управления переходами и скоростью обслуживания.** Если в системе можно управлять интенсивностями $\lambda_{ij}(u(t))$ (ускоряя восстановление из состояния S_3 или замедляя уход из состояния S_1) и (или) изменять $\mu_i(u(t))$ (выбирая более агрессивные или безопасные режимы передачи), то фактически можно контролировать долю времени π_3 (пребывание в мёртвой зоне) и эффективную скорость $\bar{\mu}$. Тем самым можно существенно понизить нагрузку ρ , уменьшить вероятность потерь и среднюю задержку. Однако на практике повышение μ_i или сокращение времени пребывания в состоянии S_3 может требовать дополнительных ресурсов (мощность, резервные каналы и т. д.), поэтому приходится искать компромисс между затратами и выигрышем в качестве обслуживания.

• **Область применения и перспективы.** Сделанные допущения (экспоненциальные распределения, марковские или полумарковские переходы) упрощают анализ и дают элегантные формулы, но могут быть лишь приближением к реальным каналам с коррелированным трафиком и сложной динамикой распространения сигнала в неоднородной среде. Тем не менее, основной вывод (необходимость выполнения условия $\lambda_{in} < \bar{\mu}$ для устойчиво-

сти) и принципы оптимизации управления $u(t)$ сохраняют силу. Предложенная модель позволяет оценить предельную производительность сети, а также адаптировать стратегию обслуживания и переходов в современных мобильных сетях, где важны как пропускная способность сети, так и энергопотребление, время восстановления связи и пр.

• **Связь с $M/G/1$ -системами с прерываниями.** Рассматриваемая в данной работе модель канала с тремя состояниями (как обобщение системы с $n > 3$ состояниями [33–35]) с изменяющейся скоростью обслуживания и периодами полной недоступности (разрывами) концептуально близка к $M/G/1$ -системам с прерываниями (см., например, работы [36, 37]). Однако, в отличие от типовых моделей $M/G/1$ с прерываниями, учитывается дополнительная управляющая функция $u(t)$, влияющая на интенсивности переходов и скорость обслуживания. Это позволяет подстраивать параметры под текущие условия мобильного канала, что расширяет классические подходы и открывает возможность адаптивной оптимизации.

• Таким образом, сходство с $M/G/1$ с прерываниями присутствует, и полученные в настоящей работе результаты (в частности, критерий $\lambda_{in} < \sum_i \pi_i \mu_i$) повторяют логику классических моделей. Однако из-за динамического управления, физической интерпретации состояний (радиоканал) и целевой функции $J(u)$ представленная модель приобретает более гибкую структуру, открывая новые возможности адаптации в инженерных приложениях.

3. ЧИСЛЕННЫЙ АНАЛИЗ УСТОЙЧИВОСТИ И ПРОИЗВОДИТЕЛЬНОСТИ АДАПТИВНОЙ МОДЕЛИ

В данном параграфе представлены результаты имитационного моделирования, подтверждающие теоретические выводы, изложенные выше. Рассматривается стохастический канал с тремя состояниями $\{S_1, S_2, S_3\}$ и входным потоком заявок интенсивности λ_{in} . Пропускная способность системы зависит от текущего состояния канала и управляющей функции $u(t)$.

Моделирование выполнено методом дискретных событий на интервале T_{max} . Исследуются ключевые характеристики: устойчивость системы, длина очереди, вероятность потерь, а также влияние параметров сети и стратегии управления на



качество работы сети. Ниже приведены параметры эксперимента, метрики и визуализация поведения системы в различных режимах.

3.1. Параметры модели

Поток заявок моделируется пуассоновским процессом с интенсивностью λ_{in} . Заявки обрабатываются по принципу FIFO (*first in, first out*) с буфером конечной ёмкости.

Канал может находиться в трёх состояниях:

- S_1 : устойчивая связь, пропускная способность μ_{norm} ;
- S_2 : ослабленный сигнал, пропускная способность μ_{slow} (или $\mu_{slow}e^{-\alpha\tau_2}$);
- S_3 : разрыв связи, обслуживание недоступно ($\mu = 0$).

Переходы между состояниями описываются либо фиксированной матрицей интенсивностей (марковский случай), либо функциями $\alpha_{ij}(t)$ в полумарковской модели. Обслуживание выполняется, если канал находится в состоянии S_1 или S_2 и очередь не пуста. В состоянии S_3 приём заявок продолжается, но обслуживание приостанавливается.

Основные метрики качества работы сети: динамика длины очереди и влияние интенсивности λ_{in} и скорости движения на устойчивость.

3.2. Результаты моделирования

Приведённые ниже графики демонстрируют поведение модели в двух режимах:

- устойчивый: $\lambda_{in} < \pi_1\mu_{norm} + \pi_2\mu_{slow}$;
- неустойчивый: $\lambda_{in} > \pi_1\mu_{norm} + \pi_2\mu_{slow}$.

В устойчивом режиме (рис. 1) очередь (размер буфера агрегатора) стабилизируется и не растёт неограниченно. Однако при перегрузке (рис. 2) имеется положительный снос.

На рис. 3 видно, что в устойчивом режиме даже при разрывах связи восстанавливается достаточно быстро. При перегрузке даже кратковременные перебои в обслуживании вызывают лавинообразный рост буфера агрегатора.

На рис. 4 показано влияние экспоненциального затухания в состоянии S_2 : при высокой нагрузке увеличивается время пребывания в состоянии S_3 , что резко ухудшает устойчивость.

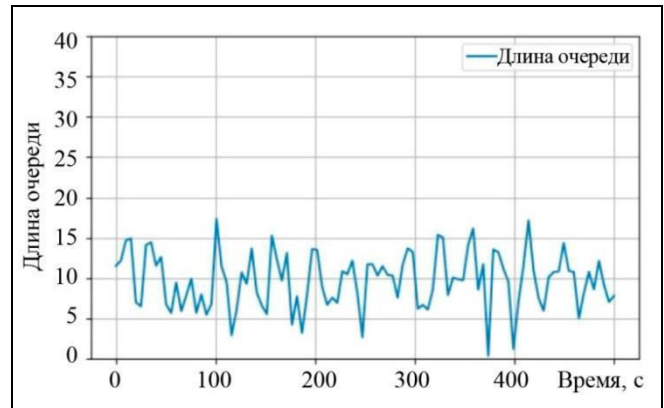


Рис. 1. Динамика длины очереди в устойчивом режиме

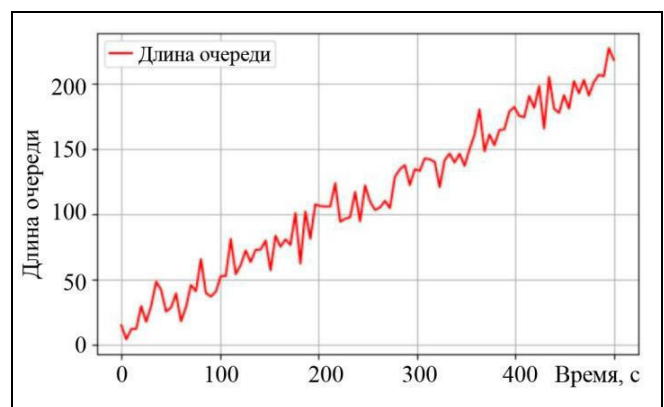


Рис. 2. Рост очереди в неустойчивом режиме

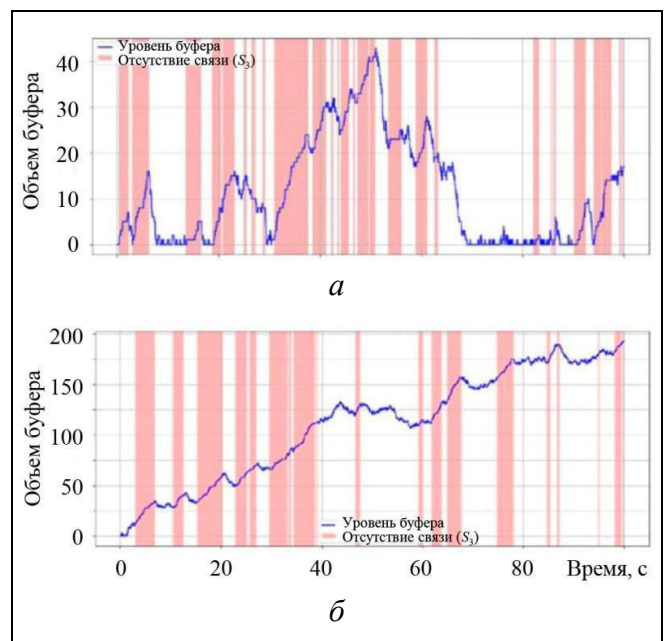


Рис. 3. Очередь: а – в устойчивом режиме ($\lambda = 3,0$ ед./с, $p = 30$ м/с), б – в неустойчивом режиме ($\lambda = 7,0$ ед./с, $p = 30$ м/с). Красные зоны – периоды разрыва связи

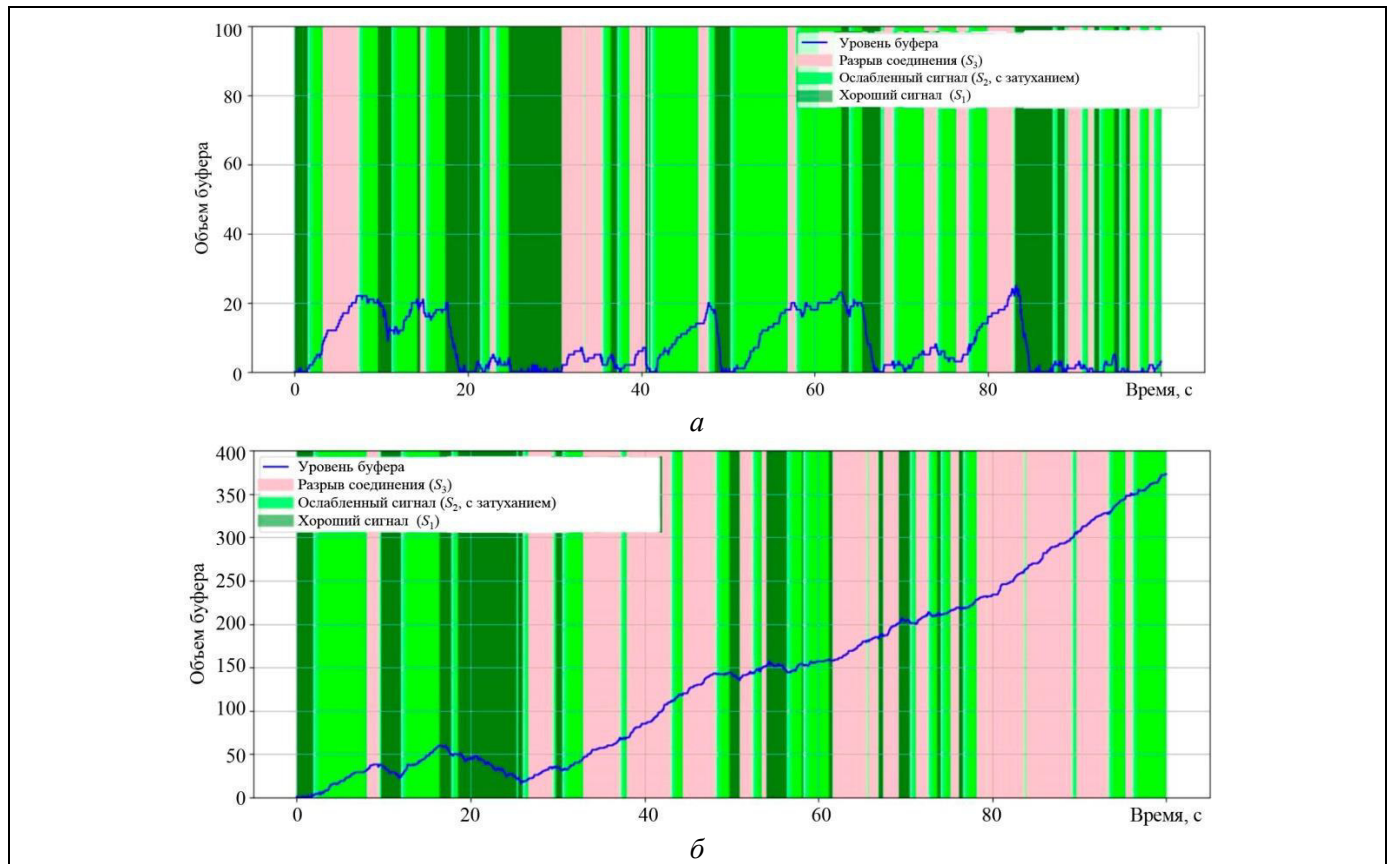


Рис. 4. Затухающая пропускная способность в состоянии S_2 : a – устойчивый режим ($\lambda = 3,0$), b – неустойчивый режим ($\lambda = 7,0$)

Рис. 5 демонстрирует преимущество адаптивного управления: вблизи критической нагрузки оно эффективно стабилизирует длину очереди.

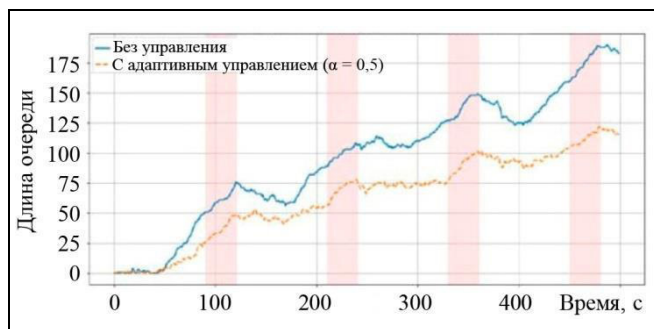


Рис. 5. Сравнение: динамика длины очереди при адаптивном управлении и при его отсутствии

На рис. 6 показана граница устойчивости. Адаптивная модель демонстрирует смещение порога перегрузки, позволяя системе функционировать ближе к критическим значениям.

Рис. 7 показывает влияние полумарковских эффектов: при гамма-распределении и распределении Вейбулла система становится более инерционной, что повышает чувствительность к перегрузкам.

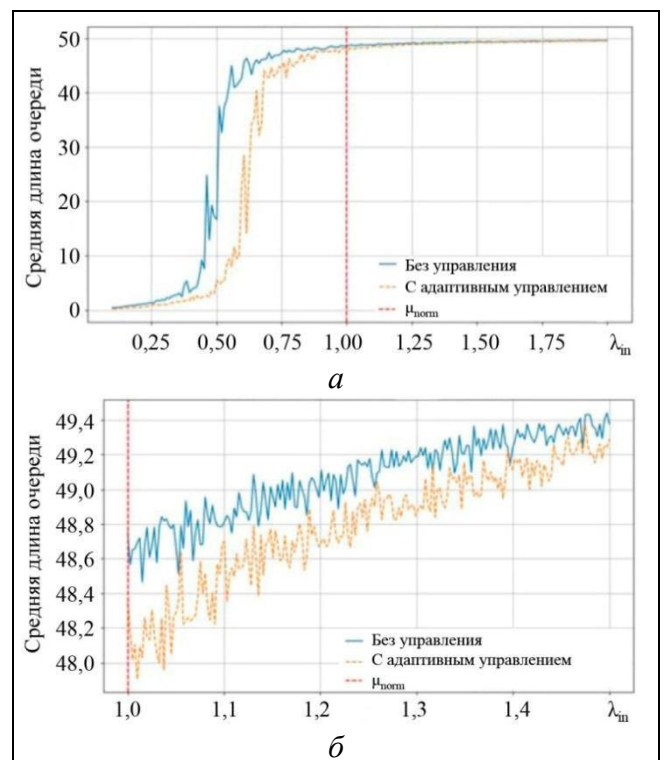


Рис. 6. Фазовый переход: средняя длина очереди в зависимости от значения λ_{in}

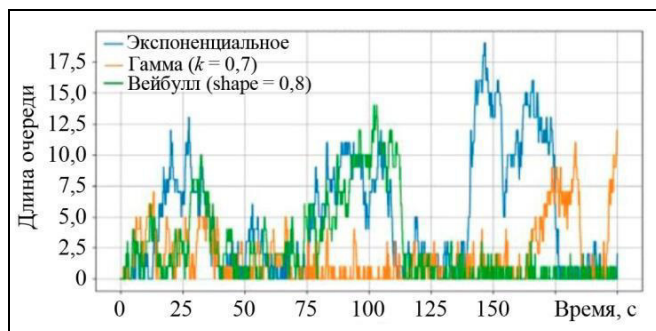


Рис. 7. Зависимость динамики длины очереди от закона распределения времени обслуживания

ЗАКЛЮЧЕНИЕ И ПЕРСПЕКТИВЫ

Предложенная модель описывает систему с адаптивным управлением в условиях стохастически меняющихся характеристик канала, находящегося в одном из трёх состояний: стабильной связи, ослабленного сигнала и полного разрыва связи. Было показано, что включение управляющей функции $u(t)$, влияющей как на интенсивность обслуживания, так и на переходы между состояниями, позволяет динамически адаптировать поведение системы к текущим условиям. В отличие от классических моделей, в которых параметры фиксированы, в предлагаемом подходе они становятся функциями стратегии, что даёт возможность прямо учитывать влияние управления на производительность сети.

Проведённый численный анализ подтверждает, что при $\lambda_{in} < \bar{\mu}$ система устойчива: очередь остаётся ограниченной. При перегрузке ($\lambda_{in} > \bar{\mu}$) происходит лавинообразный рост длины очереди. Однако адаптивное управление замедляет этот рост и фактически сдвигает границу устойчивости. Были также продемонстрированы эффекты временной корреляции: при переходе к полумарковским моделям система становится инерционной, но общие критерии устойчивости сохраняются. Это подчёркивает универсальность предложенной конструкции и её применимость как в классических, так и в более сложных телекоммуникационных сценариях.

Практическая реализация модели возможна в виде программируемого блока, встроенного в устройство агрегации. Такой модуль может использовать нейросетевые схемы для оценки текущего состояния канала и управления параметрами передачи. Кроме того, модель может быть обобщена в сторону марковского управления: когда решения зависят не только от текущего состояния, но и от времени пребывания в нём. Это естественным образом реализуется через системы с памятью

или логики, способные учитывать историю состояний. Также возможна реализация промежуточных состояний связи, задаваемых средствами многозначной логики, что позволяет описывать не только резкие, но и плавные деградации канала.

Таким образом, модель объединяет стохастическую динамику, управление и очереди в единую систему и может быть использована как в теоретических исследованиях устойчивости подобных систем, так и при разработке новых стеков протоколов передачи данных, особенно в контексте мобильных сетей нового поколения. Перспективы дальнейших исследований включают многопользовательские сценарии, исследование стратегий управления на данных и построение энергоэффективных алгоритмов для устройств с ограниченными вычислительными ресурсами на базе интеллектуального управления. В качестве самостоятельной теоретической задачи отметим исследование фазового перехода, открытым является вопрос устранения экспоненциального характера роста очереди и замедление адаптивного управления.

Автор выражает глубокую признательность анонимным рецензентам за ценные замечания и конструктивные предложения, которые значительно улучшили качество данной работы.

ЛИТЕРАТУРА

1. Патент RU 2694025 C1. Система агрегации сетевых данных в компьютерных сетях : 2019102447 : заявл. 29.02.2019 : опубл. 08.07.2019, бюл. № 19 / Марченков А.А., Есин А.А. [Patent RU 2694025 C1. Network Data Aggregation System in Computer Networks : 2019102447 : appl. 29.02.2019 : reg. 08.07.2019, bul. no. 19 / Marchenkov, A.A., Esin, A.A.]
2. Chang, C.S. Performance Guarantees in Communication Networks; Telecommunication Networks and Computer Systems. – London: Springer-Verlag, 2000. – 392 p.
3. Kleinrock, L. Queueing Systems: Volume 1. – New York: Wiley-Interscience, 1975. – 431 p.
4. Gross, D., Shortle, J., Thompson, J., Harris, C. Fundamentals of Queueing Theory; Wiley Series in Probability and Statistics. – New York: Wiley, 2011. – 527 p.
5. Baccelli, F., Blaszczyzyn, B. Stochastic Geometry and Wireless Networks: Volume I Theory // Foundations and Trends® in Networking. – 2010. – Vol. 3, no. 3–4. – P. 249–449. – DOI: <https://doi.org/10.1561/13000000006>
6. Haenggi, M., Andrews, J.G., Baccelli, F., et al. Stochastic Geometry and Random Graphs for the Analysis and Design of Wireless Networks // IEEE Journal on Selected Areas in Communications. – 2009. – Vol. 27, no. 7. – P. 1029–1046. – DOI: <https://doi.org/10.1109/JSAC.2009.090902>
7. Baccelli, F., Blaszczyzyn, B. On a Coverage Process Ranging from the Boolean Model to the Poisson–Voronoi Tessellation with Applications to Wireless Communications // Advances in Applied Probability. – 2001. – Vol. 33, no. 2. – P. 293–323. – DOI: <https://doi.org/10.1017/S0001867800010806>

8. Raigorodskii, A.M. Random Graph Models and Their Applications // Proceedings of MIPT. – 2010. – Vol. 2. – P. 130–140.
9. Stauffer, D., Aharony, A. Introduction to Percolation Theory. – Milton Park: Taylor & Francis, 1994. – 186 p.
10. Saberi, A.A. Recent Advances in Percolation Theory and Its Applications // Physics Reports. – 2015. – Vol. 578. – P. 1–32. – DOI: <https://doi.org/10.1016/j.physrep.2015.03.003>
11. Olle, H., Yuval, P., Jeffrey, E.S. Dynamical Percolation // Annales de l'Institut Henri Poincaré (B) Probability and Statistics. – 1997. – Vol. 33. – P. 497–528.
12. Li, M., Liu, R.R., Lü, L., et al. Percolation on complex networks: Theory and application // Physics Reports. – 2021. – Vol. 907. – P. 1–68. – DOI: <https://doi.org/10.1016/j.physrep.2020.12.003>
13. Badie-Modiri, A., Rizi, A.K., Karsai, M., Kivelä, M. Directed Percolation in Temporal Networks // Phys. Rev. Res. – 2022. – Vol. 4. – art. No. L022047. – DOI: <https://doi.org/10.1103/PhysRevResearch.4.L022047>
14. ElSawy, H., Zhaikhan, A., Kishk, M.A., Alouini, M.S. A Tutorial-Cum-Survey on Percolation Theory with Applications in Large-Scale Wireless Networks // IEEE Communications Surveys & Tutorials. – 2024. – Vol. 26. – P. 428–460. – DOI: <https://doi.org/10.1109/COMST.2023.3336194>
15. Lin, H., Kishk, M.A., Alouini, M.S. Performance Analysis of Infrastructure Sharing Techniques in Cellular Networks: A Percolation Theory Approach // arXiv:2502.08023. – 2025. – DOI: <https://doi.org/10.48550/arXiv.2502.08023>
16. Zhang, Q., Kassam, S.A. Finite-State Markov Model for Rayleigh Fading Channels // IEEE Transactions on Communications. – 1999. – Vol. 47, no. 11. – P. 1688–1692. DOI: 10.1109/26.803503
17. Paranthaman, V.V., Mapp, G., Shah, P., et al. Exploring Markov Models for the Allocation of Resources for Proactive Handover in a Mobile Environment // Proceedings of 2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops). – Clearwater Beach, 2015. – P. 855–861. DOI: 10.1109/LCNW.2015.7365938
18. Kalimulina, E.Y. Application of Multi-Valued Logic Models in Traffic Aggregation Problems in Mobile Networks // Proceedings of the 15th IEEE International Conference on Application of Information and Communication Technologies (AICT). – Baku, Azerbaijan, 2021. – P. 1–6. – DOI: 10.1109/AICT52784.2021.9620244
19. Kalimulina, E.Y. Lattice Structure of Some Closed Classes for Non-binary Logic and Its Applications // In: Mathematical Methods for Engineering Applications. ICMASE 2021 / ed. by F. Yilmaz, A. Queiruga-Dios, M.J. Santos Sánchez, D. Rasteiro, V. Gayoso Martínez, J. Martín Vaquero. – Springer Proceedings in Mathematics & Statistics. – Vol. 384. – Cham: Springer, 2022. – P. 25–34. – DOI: 10.1007/978-3-030-96401-6_2
20. Goldsmith, A. Wireless Communications. – Cambridge: Cambridge University Press, 2005. – 648 p.
21. Semi-Markov Models and Applications / ed. by J. Janssen, N. Limnios. – Cham: Springer Science & Business Media. – 2013.
22. Limnios, N., Oprisan, G. Semi-Markov Processes and Reliability. – Cham: Springer Science & Business Media. – 2012.
23. Cinlar, E. Introduction to Stochastic Processes. – Hoboken: Prentice-Hall, 1975.
24. Ross, S. Stochastic Processes. – New York: Wiley, 1996. – 328 p.
25. Kirk, D.E. Optimal Control Theory: An Introduction. – Hoboken: Prentice-Hall, 1970.
26. Bertsekas, D.P. Dynamic Programming and Optimal Control. – Raleigh: Athena Scientific, 2005. – 484 p.
27. Boyd, S., Vandenberghe, L. Convex Optimization. – Cambridge: Cambridge University Press, 2004.
28. Wang, T., Proakis, J.G., Masry, E., Zeidler, J.R. Performance Degradation of OFDM Systems due to Doppler Spreading // IEEE Transactions on Wireless Communications. – 2006. – Vol. 5, no. 6. – P. 1422–1432. – DOI: 10.1109/TWC.2006.1638663
29. Yin, X., Cheng, X. Propagation Channel Characterization, Parameter Estimation, and Modeling for Wireless Communications. – Hoboken: John Wiley & Sons, 2016. – 350 p.
30. Modeas, I., Kaloxylas, A., Merakos, L., Tsolkas, D. An Adaptive and Distributed Network Selection Mechanism for 5G Networks // Computer Networks. – 2021. – Vol. 189. – Art. no. 107943. – DOI: 10.1016/j.comnet.2021.107943
31. Kalimulina, E.Y. Mathematical Model for Reliability Optimization of Distributed Telecommunications Networks // Proceedings of 2011 International Conference on Computer Science and Network Technology. – Harbin, China, 2011. – P. 2847–2853. – DOI: <https://doi.org/10.1109/ICCSNT.2011.61825570>
32. Kalimulina, E.Y. A New Approach for Dependability Planning of Network Systems // International Journal of System Assurance Engineering and Management. – 2013. – Vol. 4. – P. 215–222.
33. Есин А.А. Анализ современных принципов построения систем управления на основе моделей многозначной логики // Управление большими системами. – 2020. – Вып. 88. – С. 69–98. – DOI: <https://doi.org/10.25728/ubs.2020.88.4>. [Esin, A.A. Analysis and Design Principles of Modern Control Systems Based on Multi-valued Logic Models // Large-Scale Systems Control. – 2020. – Iss. 88. – P. 69–98. – DOI: <https://doi.org/10.25728/ubs.2020.88.4>. (In Russian)]
34. Esin, A.A. Characteristics of Structurally Finite Classes of Order-Preserving Three-Valued Logic Maps // Logic Journal of the IGPL. – 2024. – Art. no. jzae128. – URL: <https://academic.oup.com/jigpal/article-lookup/doi/10.1093/jigpal/jzae128>.
35. Esin, A.A. Structural Analysis of Precomplete Classes and Closure Diagrams in Multi-valued Logic // Iranian Journal of Fuzzy Systems. – 2025. – Vol. 21, no. 6. – P. 127–145.
36. Gao, S., Liu, Z. An M/G/1 Queue with Single Working Vacation and Vacation Interruption under Bernoulli Schedule // Applied Mathematical Modelling. – 2013. – Vol. 37, no. 3. – P. 1564–1579. – DOI: 10.1016/j.apm.2012.04.045
37. Zhang, M., Hou, Z. Performance Analysis of M/G/1 Queue with Working Vacations and Vacation Interruption // Journal of Computational and Applied Mathematics. – 2010. – Vol. 234, no. 10. – P. 2977–2985. – DOI: 10.1016/j.cam.2010.04.010

Статья представлена к публикации членом редколлегии
В. М. Вишневым.

Поступила в редакцию 17.05.2025,
после доработки 19.05.2025.
Принята к публикации 20.05.2025.

Есин Антон Анатольевич – директор по развитию,
АО «МирВайФай», г. Москва, ✉ anton.esin@imm.am.
ORCID iD: <https://orcid.org/0009-0007-0680-8288>

© 2025 г. Есин А. А.



Эта статья доступна по лицензии Creative Commons
«Attribution» («Атрибуция») 4.0 Всемирная.



A MATHEMATICAL MODEL OF ADAPTIVE TRAFFIC CONTROL IN MOBILE NETWORKS WITH VARIABLE SIGNAL QUALITY

A. A. Esin

AO MirWiFi, Moscow, Russia

✉ anton.esin@imm.am

Abstract. This paper considers the mathematical modeling problem of traffic transmission in mobile networks under high user mobility and spatially heterogeneous coverage, including signal degradation (“dead”) zones. Traffic aggregation at the channel level is applied to increase the reliability and stability of data transmission. A Markov model of a communication channel is proposed to study the effectiveness of aggregation algorithms and adapt them to network operating parameters and user velocity. The model is based on a periodic affine motion of a mobile device between base stations (BSs) uniformly distributed along a straight line. Within this model, the concepts of stable coverage zones and transition zones are introduced in terms of the distances to the nearest and next BSs. The channel state is described by a Markov chain with states corresponding to signal quality sampling: stable connection, degraded connection, and disconnection. Transitions between states are governed by a continuous-time Markov process with constant rates, and the parameters of this process are determined from empirical network data. An extension of the model to incorporate the time dependence of channel states is also considered, leading to a semi-Markov framework. For both cases (Markov and semi-Markov chains), explicit expressions are derived for the stationary probabilities of states, and system stability conditions are formulated to ensure bounded traffic queues. In addition, an adaptive control model of the channel throughput is proposed; this model optimizes transmission parameters depending on the current channel state, request queue length, and user velocity. The effectiveness of the approach is demonstrated by numerical simulations: the network has stable performance across a wide range of mobility levels and coverage parameters. The model can be applied to the reliability analysis and optimization of network protocols in highly mobile environments, including high-speed railway transport, vehicle networks, and mobile platforms.

Keywords: mobile networks, Markov model, semi-Markov process, system stability, Laplace transform.

Acknowledgments. The author is grateful to the reviewers for careful reading of the manuscript and helpful remarks.