

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ



НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ

САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО
ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Информатика. Телекоммуникации.
Управление

ТОМ 10, № 3
2017

Издательство Политехнического университета
Санкт-Петербург
2017

НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ САНКТ-ПЕТЕРБУРГСКОГО
ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА
ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ

РЕДАКЦИОННЫЙ СОВЕТ ЖУРНАЛА

Председатель

Юсупов Р.М., чл.-кор. РАН;

Редакционный совет:

Абрамов С.М., чл.-кор. РАН;

Арсеньев Д.Г., д-р техн. наук, профессор;

Воеводин В.В., чл.-кор. РАН;

Заборовский В.С., д-р техн. наук, профессор;

Козлов В.Н., д-р техн. наук, профессор;

Фотиади А.Э., д-р физ.-мат. наук, профессор;

Черноруцкий И.Г., д-р техн. наук, профессор.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА

Главный редактор

Коротков А.С., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Редакционная коллегия:

Бабкин А.В., д-р экон. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Ицыксон В.М., канд. техн. наук, доцент, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Prof. Dr. Philippe Ferrari, Head of the RF and Millimeter-Wave Lab IMEP-LAHC Microelectronics, Electromagnetism and Photonic Institute, Grenoble Alpes University, France;

Клавдиев В.Е., канд. техн. наук, доцент, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Prof. Dr. Wolfgang Krautschneider, Head of Nanoelectronics Institute, Hamburg University of Technology, Germany;

Кучерявый Е.А., канд. техн. наук, профессор, Tampere University of Technology, Finland.

Dr. Fa-Long Luo, Chief Scientist, Element CXI, San Jose, USA;

Макаров С.Б., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Prof. Dr. Emil Novakov, IMEP-LAHC Microelectronics, Electromagnetism and Photonic Institute, Grenoble, France;

Трифонов П.В., канд. техн. наук, доцент, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Устинов С.М., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Цикин И.А., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия;

Шкодырев В.П., д-р техн. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого, Россия.

Журнал с 1995 года издается под научно-методическим руководством Российской академии наук. С 2008 года выпускается в составе сериального периодического издания «Научно-технические ведомости СПбГПУ» ISSN 1994-2354.

Журнал с 2002 года входит в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

Сведения о публикациях представлены в Реферативном журнале ВИНТИ РАН, в международной справочной системе «Ulrich`s Periodical Directory», в базах данных Российский индекс научного цитирования (РИНЦ), Google Scholar, EBSCO, Math-Net.Ru, ProQuest, Index Copernicus

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

При перепечатке материалов ссылка на журнал обязательна.

Точка зрения редакции может не совпадать с мнением авторов статей.

Адрес редакции и издательства: Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.
Тел. редакции (812) 552-62-16.

THE MINISTRY OF EDUCATION AND SCIENCE OF THE RUSSIAN FEDERATION



**ST. PETERSBURG STATE
POLYTECHNICAL UNIVERSITY
JOURNAL**

Computer Science.
Telecommunications and Control Systems

**Vol. 10, No. 3
2017**

Polytechnical University Publishing House
Saint Petersburg
2017

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS

EDITORIAL COUNCIL

Head of the editorial council

Prof. Dr. *Rafael M. Yusupov* (corresponding member of the Russian Academy of Sciences)

Members:

Prof. Dr. *Sergey M. Abramov* (corresponding member of the Russian Academy of Sciences),

Prof. Dr. *Dmitry G. Arseniev*,

Prof. Dr. *Vladimir V. Voevodin* (corresponding member of the Russian Academy of Sciences),

Prof. Dr. *Vladimir S. Zaborovsky*,

Prof. Dr. *Vladimir N. Kozlov*,

Prof. Dr. *Alexandr E. Fotiadi*,

Prof. Dr. *Igor G. Chernorutsky*.

EDITORIAL BOARD

Editor-in-chief

Prof. Dr. *Alexander S. Korotkov*, Peter the Great St. Petersburg Polytechnic University, Russia;

Members:

Prof. Dr. *Alexandr V. Babkin*, Peter the Great St. Petersburg Polytechnic University, Russia;

Assoc. Prof. Dr. *Vladimir M. Itsyson*, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Philippe Ferrari*, Head of the RF and Millimeter-Wave Lab IMEP-LAHC Microelectronics, Electromagnetism and Photonic Institute, Grenoble Alpes University, France;

Assoc. Prof. Dr. *Vladimir E. Klavdiev*, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Yevgeni Koucheryavy*, Tampere University of Technology, Finland.

Prof. Dr. *Wolfgang Krautschneider*, Head of Nanoelectronics Institute, Hamburg University of Technology, Germany;

Dr. *Fa-Long Luo*, Chief Scientist, Element CXI, San Jose, USA;

Prof. Dr. *Sergey B. Makarov*, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Emil Novakov*, IMEP-LAHC Microelectronics, Electromagnetism and Photonic Institute, Grenoble, France;

Prof. Dr. *Viacheslav P. Shkodyrev*, Peter the Great St. Petersburg Polytechnic University, Russia;

Assoc. Prof. Dr. *Peter V. Trifonov*, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Igor A. Tsikin*, Professor, Peter the Great St. Petersburg Polytechnic University, Russia;

Prof. Dr. *Sergey M. Ustinov*, Peter the Great St. Petersburg Polytechnic University, Russia.

The journal is published under scientific and methodical guidance of the Russian Academy of Sciences since 1995. The journal is published since 2008 as part of the periodical edition «Nauchno-tekhnicheskie vedomosti SPbGPU» (ISSN 1994-2354).

The journal is included in the List of Leading Peer-Reviewed Scientific Journals and other editions to publish major findings of PhD theses for the research degrees of Doctor of Sciences and Candidate of Sciences.

The journal is indexed by Ulrich's Periodicals Directory, Google Scholar, EBSCO, ProQuest, Index Copernicus, VINITI RAS Abstract Journal (Referativnyi Zhurnal), VINITI RAS Scientific and Technical Literature Collection, Russian Science Citation Index (RSCI) database © Scientific Electronic Library and Math-Net.ru databases.

The journal is registered with the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR). Certificate ПИ № ФС77-51457 issued Oct. 19, 2012.

No part of this publication may be reproduced without clear reference to the source.

The views of the authors can contradict the views of the Editorial Board.

The address: 195251 Polytekhnikeskaya Str. 29, St. Petersburg, Russia.

Содержание

Телекоммуникационные системы и компьютерные сети

Попов С.Г., Тучков А.С., Курочкин Л.М., Шарагин М.П. Алгоритм распределения потоков данных между беспроводными глобальными и локальными сетями связи	7
Пастушок И.А., Тюрликов А.М. Анализ эффективности использования беспроводных централизованных сетей для передачи неадаптивного видео по протоколу HTTP	18
Дробинцев П.Д., Котляров В.П., Левченко А.В. Экспериментальные аспекты оценки пропускной способности памяти крупномасштабных систем с архитектурой ccNUMA	32

Устройства и системы передачи, приема и обработки сигналов

Крнета М., Пятак И.М. Разработка цифровой коррекции характеристик АЦП для приемника прямого усиления на основе ПЛИС	44
Ахметов Д.Б., Аль-Кархи Омар Дж. Генератор, управляемый цифровым кодом, с пониженным энергопотреблением на основе 65 нм КМОП-технологии	53

Программное обеспечение вычислительных, телекоммуникационных и управляющих систем

Березун Д.А. Полная головная линейная редукция	59
---	----

Моделирование вычислительных, телекоммуникационных, управляющих и социально-экономических систем

Смирнов А.С., Смольников Б.А., Леонтьев В.А. Принципы неголономной механики в управлении движением группы транспортных роботов	83
Чаймарданов П.А., Былина М.С. Новая методика расчета оптического усилителя EDFA с многоканальными источниками сигналов и накачек	92
Горковенко Д.К. Сравнительный анализ моделей эпидемии и клеточного автомата при моделировании распространения информации в социальных сетях	103

Contents

Telecommunications Systems and Computer Networks

Popov S.G., Tuchkov A.S., Kurochkin L.M., Sharagin M.P. <i>Algorithm of Data Streams Distribution between Wireless WAN and LAN</i>	7
Pastushok I.A., Turlikov A.M. <i>The Performance Analysis of Centralized Wireless Networks for Non-Adaptive HTTP-based Video Transmission</i>	18
Drobintsev P.D., Kotlyarov V.P., Levchenko A.V. <i>Experimental Aspects of Memory Bandwidth for HPC Systems with ccNUMA Architecture</i>	32

Circuits and Systems for Receiving, Transmitting and Signal Processing

Krneta M., Piatak I.M. <i>Linearization of the Analog-to-Digital Converter for an FPGA-based Direct Digital Receiver</i>	44
Akhmetov D.B., Al-Karkhi Omar J. <i>A Low-Power Digitally Controlled Oscillator Based on 65-nm CMOS Technology</i>	53

Software of Computer, Telecommunications and Control Systems

Berezun D.A. <i>Complete Head Linear Reduction</i>	59
---	----

Simulations of Computer, Telecommunications, Control and Social Systems

Smirnov A.S., Smolnikov B.A., Leontev V.A. <i>Principles of Nonholonomic Mechanics in Controlling a Transport Robot Group</i>	83
Chaymardanov P.A., Bylina M.S. <i>A Novel Calculation Method for Optical Amplifier EDFA with Multichannel Signal and Pump Sources</i>	92
Gorkovenko D.K. <i>Comparison of Epidemic Models and Cellular Automata in Modeling of Diffusion of Information in Social</i>	103



Телекоммуникационные системы и компьютерные сети

DOI: 10.18721/JCSTCS.10301

УДК 004.22, 658.5, 65.011.56

АЛГОРИТМ РАСПРЕДЕЛЕНИЯ ПОТОКОВ ДАННЫХ МЕЖДУ БЕСПРОВОДНЫМИ ГЛОБАЛЬНЫМИ И ЛОКАЛЬНЫМИ СЕТЯМИ СВЯЗИ

С.Г. Попов, А.С. Тучков, Л.М. Курочкин, М.П. Шарагин

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

Описан алгоритм динамического распределения потоков данных между беспроводными локальными и глобальными сетями связи, такими как Wi-Fi и LTE. Предмет исследования – зависимость времени выполнения алгоритма распределения потоков данных между беспроводными локальными и глобальными сетями связи. Разработан алгоритм динамического мультиплексирования гетерогенных каналов связи, учитывающий динамику оценки пропускных способностей каналов связи и их комбинаций. Проведено исследование времени работы алгоритма выбора канала, показавшее эффективность стратегии безусловного приоритета перед стратегией квотирования. Реализованный прототип системы может использоваться в подсистеме маршрутизации мультипротокового узла.

Ключевые слова: алгоритм распределения потоков данных; мультипротокольный узел; интеллектуальная транспортная система; гетерогенные каналы связи; беспроводные технологии; динамическое мультиплексирование каналов.

Ссылка при цитировании: Попов С.Г., Тучков А.С., Курочкин Л.М., Шарагин М.П. Алгоритм распределения потоков данных между беспроводными глобальными и локальными сетями связи // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2017. Т. 10. № 3. С. 7–17. DOI: 10.18721/JCSTCS.10301

ALGORITHM OF DATA STREAMS DISTRIBUTION BETWEEN WIRELESS WAN AND LAN

S.G. Popov, A.S. Tuchkov, L.M. Kurochkin, M.P. Sharagin

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

The goal of the study is the development and experimental verification of the applicability of a new data streams distribution algorithm between wireless local and global communication networks. The object of research is the data streams dynamic allocation algorithm between the wireless local and wide area networks such as Wi-Fi and LTE. The research subject is the dependent runtime data flow allocation algorithm between wireless local and wide area networks. The authors developed a dynamic multiplexing algorithm for heterogeneous communication channels that takes into account the dynamics of evaluating the communication channels capacity and channel combinations, investigated the running time of the channel selection algorithm, which showed that the unconditional priority strategy was more effective than the quota strategy. The implemented prototype

system can be used in a multiprotocol routing subsystem.

Keywords: data streams distribution algorithm; multiprotocol unit; intelligent transports system; heterogeneous communication channels; wireless technology; dynamic multiplexing.

Citation: Popov S.G., Tuchkov A.S., Kurochkin L.M., Sharagin M.P. Algorithm of data streams distribution between wireless WAN and LAN. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems. 2017, Vol. 10, No. 3, Pp. 7–17. DOI: 10.18721/JCSTCS.10301

Введение

Пропускная способность трафика, формируемого сервисами Интернет и системными службами, используемого в мультипротокольном узле [7], превышает пропускную способность используемых каналов передачи данных, что приводит к рассогласованию функционирования компонент интеллектуальной транспортной системы (ИТС). Поэтому при внедрении ИТС возникает проблема выбора такой комбинации каналов связи, которая обеспечит минимальную гарантированную пропускную способность для потока данных и при этом минимизирует время передачи.

Основная концепция ИТС заключается в одновременном доступе к нескольким сетям, однако каналы связи используются изолированно друг от друга, что снижает эффективность их использования. Эффективным подходом к решению проблемы полного использования суммарной про-

пускной способности гетерогенных каналов является мультиплексирование, что обеспечивает непрерывную передачу данных адресату по нескольким каналам связи.

В работах [2, 3] авторы решали проблему повышения эффективности передачи потоков данных по разнородным каналам передачи данных с использованием протокола МРТСР. Для решения проблемы превышения потока данных пропускной способности передачи данных в транспортных сетях в работах [1, 4] использованы гомогенные каналы передачи данных.

В статьях [5, 6] предложен новый способ распределения трафика в узле: концепция совместного роутинга, основанная на оценке задержки и числа потоков отдельно для внутренних и внешних соединений, основанных на учёте дистанции между транспортными средствами.

В статье [10] целью исследования являлось определение времени соединения и скорости передачи данных из транспортно-



Рис. 1. Классификация технологий распределения потоков данных

го средства в стационарную беспроводную сеть, используя протоколы FTP, HTTP, HTTPS и SSH. Исследование базировалось на стандарте беспроводных сетей IEEE 802.11g.

Анализ данных работ позволил сформировать классификацию технологий распределения данных (рис. 1), основанную на представлении технологий относительно технических возможностей реализации мультипротокольности и способа оценки пропускной способности каналов связи. Существующие технологии оценивают пропускные способности каналов статически, а поток данных обрабатывается только от одного приложения, что требует разработки нового алгоритма динамического мультиплексирования гетерогенных каналов передачи данных.

Наша цель – разработка и экспериментальная проверка области применимости

нового алгоритма распределения потоков данных между беспроводными локальными и глобальными сетями связи.

Мультипротокольная среда передачи данных транспортных средств

На рис. 2 приведена программно-аппаратная схема фрагмента стенда мультипротокольного узла. Анализатор трафика (Traffic analyzer) обрабатывает телекоммуникационный трафик, поступающий от приложений, установленных на мультипротокольном узле. К ним могут относиться системы передачи голосовой связи, системы мультимедийного контента, электронная почта, компоненты систем безопасности, информация о дорожных условиях и другие. Анализатор разбирает входящий трафик на заявки, каждая из которых имеет свой приоритет, требуемую пропускную способность канала для передачи данных и

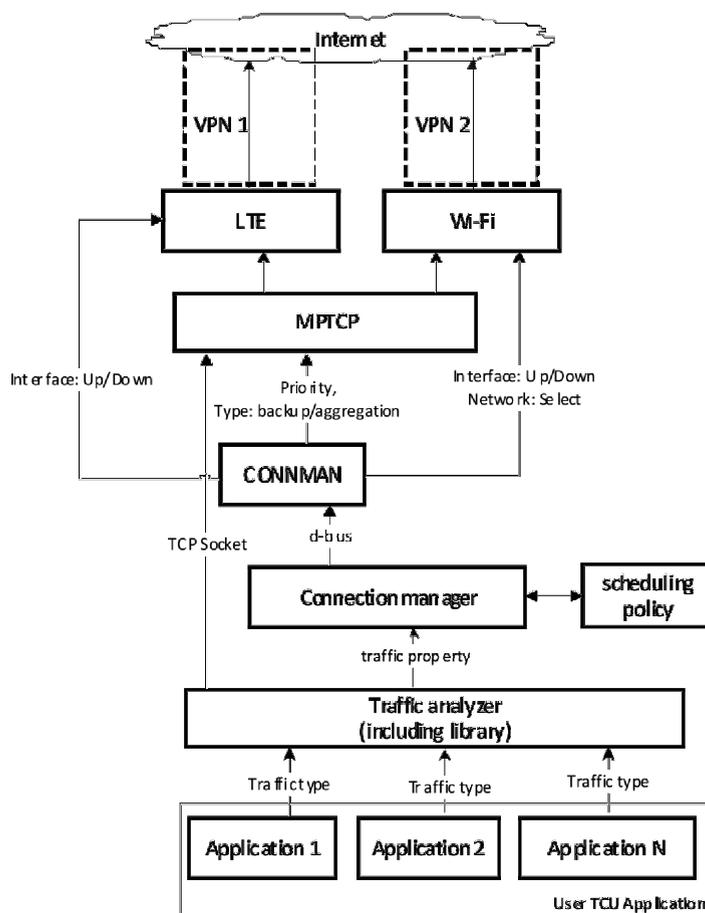


Рис. 2. Программно-аппаратная схема фрагмента стенда мультипротокольного узла

объем передаваемых данных.

Диспетчер соединений (**Connection manager**) контролирует все сетевые подключения по запросам от приложений, а также формирует пары «заявка–канал передачи данных». Обработка заявок регламентируется правилами, заложенными планировщиком (**Scheduling policy**), которые могут со временем изменяться.

Обработанный телекоммуникационный трафик передается по разнородным беспроводным каналам связи по протоколу МРТСП [8, 9], что позволяет обмениваться пакетами с несколькими адресами/интерфейсами одновременно, в рамках одного соединения. Отметим, что МРТСП сохраняет обратную совместимость со старыми версиями ТСП и допускает подключение устройства по стандартному ТСП.

Для передачи данных используется технология мультиплексирования, позволяющая динамически формировать и одновременно использовать виртуальные каналы связи мультипротокольного узла, с целью ускорения передачи данных (наращивания пропускной способности). Виртуальный канал передачи данных может состоять из нескольких каналов, физически установленных на мультипротокольном узле. Его суммарная пропускная способность равна экспериментальной оценке пропускной способности одновременного использования физических каналов.

Критерий оптимальности задачи распределения потоков данных

Пусть $F(t_1...t_n)$ – множество заявок для обслуживания, формирующих входящий поток данных. Каждая t_i заявка характеризуется следующими параметрами: приоритет p_j , где $j = \overline{1...3}$, требуемая пропускная способность для обслуживания заявки C_{t_i} , объем передаваемых данных V_{t_i} .

В стационарном случае суммарная пропускная способность виртуального канала равна сумме пропускных способностей входящих в него каналов:

$$C = \sum C_i, \quad (1)$$

где C_i – пропускная способность i -го фи-

зического канала, заявленная в спецификации используемой технологии передачи данных.

Учет динамики текущей пропускной способности виртуальных каналов связи обеспечивает оценку пропускной способности канала на текущий момент времени (более точный выбор канала для передачи данных):

$$C = \sum \hat{C}_i, \quad (2)$$

где \hat{C}_i – динамическая оценка пропускной способности i -го виртуального канала передачи данных.

На практике суммарная пропускная способность виртуального канала меньше, чем теоретически возможная. Это связано с накладными расходами на создание соединения между источником и приемником данных. Накладные расходы зависят от времени создания подпотоков передачи данных и их разбиения, количества каналов, разнородности каналов, качества передачи пакетов в канале (процент потерь пакетов) и задержек в сети. За разбиение и создание подпотоков отвечает специальный планировщик потоков протокола МРТСП. За качество разбиения отвечает алгоритм распределения нагрузки, от реализации которого зависит разница между теоретическим и эмпирическим значениями пропускной способности виртуального канала. Иными словами, при учете динамики пропускной способности необходимо ввести коэффициент пропускной способности виртуального канала связи. Тогда формула (2) примет следующий вид:

$$C = k_i \sum \hat{C}_i, \quad (3)$$

где k_i – коэффициент пропускной способности i -го виртуального канала, сформированного для передачи по протоколу МРТСП ($k_i < 1$).

Во время движения транспортного средства количество доступных телекоммуникационных сетей меняется, что требует непрерывного обновления списка комбинаций каналов связи.

Таким образом, осуществляется поиск

локального минимума задачи двухкритериальной оптимизации, представленной в следующем виде:

$$\arg \min\{t, \max(k_i \sum \hat{C}_i)\}, \quad (4)$$

где t — время передачи потока данных.

Алгоритм динамического мультиплексирования гетерогенных каналов передачи данных

Алгоритм динамического мультиплексирования гетерогенных каналов передачи данных состоит из следующих процедур: формирования пула заявок, распределения заявок по каналам связи, расчёта новых комбинаций каналов и оценки пропускных способностей каналов.

Алгоритм выполняется бесконечное число раз, пока функционирует транспортное средство в ИТС. Процедура расчёта новых комбинаций каналов периодически опрашивает активные на транспортном средстве интерфейсы передачи данных, составляет из них всевозможные комбинации виртуальных каналов полным перебором и сохраняет их в память. После того как были проинициализированы новые каналы связи, алгоритм оценивает пропускную

способность каждого из каналов и обновляет матрицу пропускных способностей.

Поскольку одновременно установленных каналов передачи данных на мультипротокольном узле не может быть много (обычно не более 6), то за полиномиальное время можно спрогнозировать комбинации виртуальных каналов, формируемых как комбинации физических каналов связи.

Алгоритм формирования пула заявок. Очередь и пул заявок имеют фиксированный размер, причем размер очереди заявок больше размера пула. Это обусловлено тем, что количество виртуальных каналов значительно меньше числа поступающих в очередь заявок.

Блок-схема алгоритма приведена на рис. 3. При поступлении новой заявки t_i происходит проверка на переполнение очереди. В случае если очередь не переполнена, происходит поиск канала среди всех свободных каналов на текущий момент времени, удовлетворяющего минимальной пропускной способности заявки по критерию $\min(\hat{C}_k - C_{t_i})$, где \hat{C}_k — текущая оценка k -го виртуального канала передачи данных. Таким образом алгоритм ищет канал связи, у которого разность между его пропускной способностью и пропускной



Рис. 3. Блок-схема алгоритма формирования пула заявок

способностью заявки минимальна. После поиска канала осуществляется проверка переполнения пула. В случае если пул не переполнен, заявка добавляется в пул и удаляется из очереди. Пул представляет собой словарь, хранящий пары «id заявки – виртуальный канал».

Алгоритм распределения заявок по каналам связи. Сортировка пула осуществляется по времени передачи заявок $t_{s_i} = \frac{V_{t_i}}{C_{t_i}}$. Процедура сортировки инициализируется в момент добавления новой заявки в пул.

После того как пул отсортирован, заявка отправляется на передачу по виртуальному каналу. Перед отправкой каналу присваивается статус «занят». После отправления заявки канал получает статус «свободен», позволяющий вновь использовать данный канал в алгоритме формирования пула. Заявка после отправления на передачу удаляется из пула. За гарантию доставки заявки отвечает протокол МРТСР.

В случае если заявку не удалось отправить на передачу, вновь вызывается процедура сортировки пула, т. к. в этот момент могла появиться новая заявка, время передачи которой меньше, чем у той, которую не удалось отправить. Блок-схема алгоритма распределения заявок по каналам связи приведена на рис. 4.

Алгоритм распределения потоков данных инвариантен к стратегиям обслуживания очереди: они являются входными данными для алгоритма. Данный алгоритм можно использовать для технологической среды тестирования стратегий обслуживания очереди и пула заявок.

Условия проведения экспериментов

Цель экспериментов – обоснование и проверка области применимости алгоритма динамического распределения потоков данных между беспроводными локальными и глобальными сетями связи. Для проведения экспериментов были рассмотрены три приоритета телекоммуникационного трафика: высший приоритет имеют сооб-



Рис. 4. Блок-схема алгоритма распределения заявок по каналам связи

щения «emergency call», средний приоритет имеет трафик от мультимедийных приложений, электронной почты, приложений, отражающих ситуацию на дороге и пр., низший приоритет имеет трафик от фоновых процессов, происходящих на мультипротокольном узле.

Для подтверждения применимости алгоритма динамического мультиплексирования гетерогенных каналов связи проведены два блока экспериментов:

первый блок состоит из четырех серий экспериментов, цель которых – определение зависимости времени работы алгоритмов формирования пула и распределения заявок по каналам от размера очереди и пула (табл. 1);

второй блок состоит также из четырех серий экспериментов, цель которых – определение распределения времени работы алгоритмов формирования пула и распределения заявок по каналам на минимальных и максимальных значениях размеров очереди и пула (табл. 2).

Результаты экспериментов

В ходе проведения первого блока экспериментов получены серии значений

Таблица 1

Описание параметров первого блока экспериментов

Параметры	Эксперимент			
	1	2	3	4
Размер структуры, шт.	от 8 до 106	от 8 до 106	от 6 до 104	от 6 до 104
Эмулируемые каналы, шт.	2 x Wi-Fi, 2 x LTE			
Стратегия обслуживания очереди	Безусловный приоритет	По квотам	Безусловный приоритет	По квотам
Распределение заявок по приоритетам	(10;45;45)	(10;45;45)	(10;45;45)	(10;45;45)
Размер потока, шт.	1000	1000	1000	1000
Частота ПЗ, Гц	2	2	2	2

Таблица 2

Описание параметров второго блока экспериментов

Параметры	Эксперимент			
	1	2	3	4
Размер очереди, шт.	8	8	106	106
Размер пула, шт.	6	6	104	104
Эмулируемые каналы, шт	2 x Wi-Fi, 2 x LTE			
Стратегия обслуживания очереди	Безусловный приоритет	По квотам	Безусловный приоритет	По квотам
Распределение заявок по приоритетам	(10;45;45)	(10;45;45)	(10;45;45)	(10;45;45)
Размер потока, шт.	1000	1000	1000	1000
Частота ПЗ, Гц	2	2	2	2

среднего времени обслуживания заявок для алгоритмов формирования пула и распределения заявок по каналам в зависимости от выбранных правил формирования пула.

На рис. 5 представлена зависимость среднего времени обслуживания заявки в очереди от размера очереди. Серия экспериментов показала, что среднее время обслуживания заявки не зависит от размеров очереди и пула, и оно константно. Также среднее время обслуживания заявки в очереди для алгоритма формирования пула для стратегии безусловного приоритета меньше, чем для стратегии «по квотам».

На рис. 6 представлена зависимость среднего времени распределения заявки в пуле. Серия экспериментов показала, что среднее время обслуживания заявки в пуле для алгоритма распределения заявок по каналам передачи данных не зависит от выбранной стратегии формирования пула. Связано это с тем, что в стратегии безусловного приоритета происходит только одна процедура сортировки очереди заявок, в то время как в стратегии «по квотам» необходимо сортировать заявки трижды – наборы заявок по приоритетам.

В ходе проведения второго блока экс-

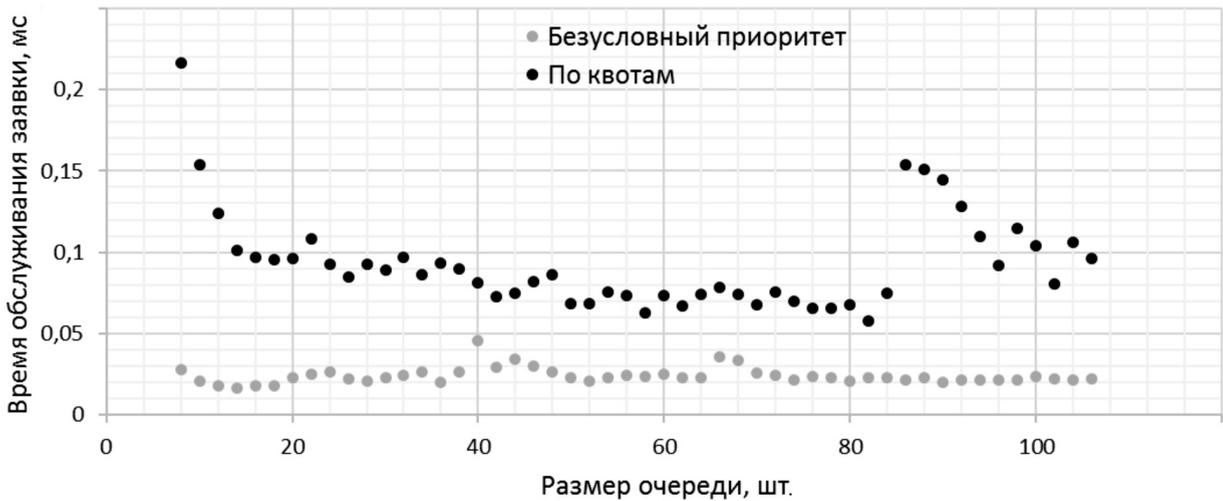


Рис. 5. Зависимость среднего времени обслуживания заявки в очереди от размера очереди

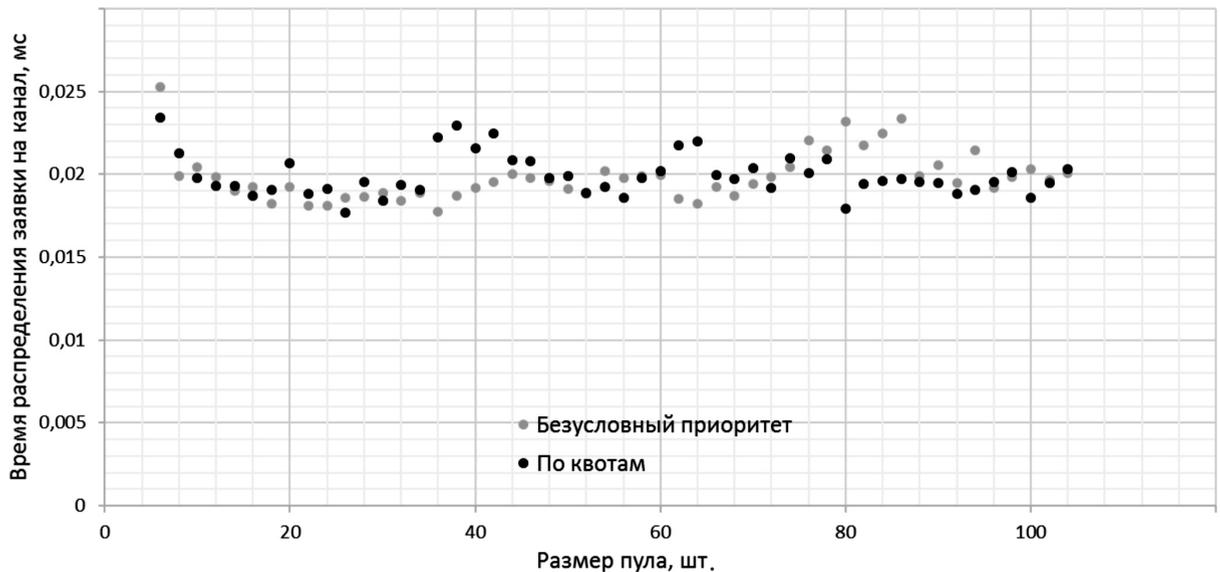


Рис. 6. Зависимость среднего времени распределения заявки в пуле

периментов получены значения среднего времени работы обслуживания заявок для алгоритмов формирования пула и распределения заявок по каналам в зависимости от выбранных правил формирования пула на максимальных и минимальных значениях пула и очереди заявок.

На рис. 7 а представлены гистограммы распределения среднего времени обслуживания заявок на минимальных и максимальных значениях размера очереди для

алгоритма формирования пула на 20 интервалах. Из гистограмм видно, что среднее время обслуживания заявок в очереди на максимальных и минимальных значениях размеров очередей для обеих стратегий не отличается. Также видно, что среднее время обслуживания заявок в очереди для стратегии безусловного приоритета меньше, чем для стратегии «по квотам», что подтверждается результатами из первого блока экспериментов.

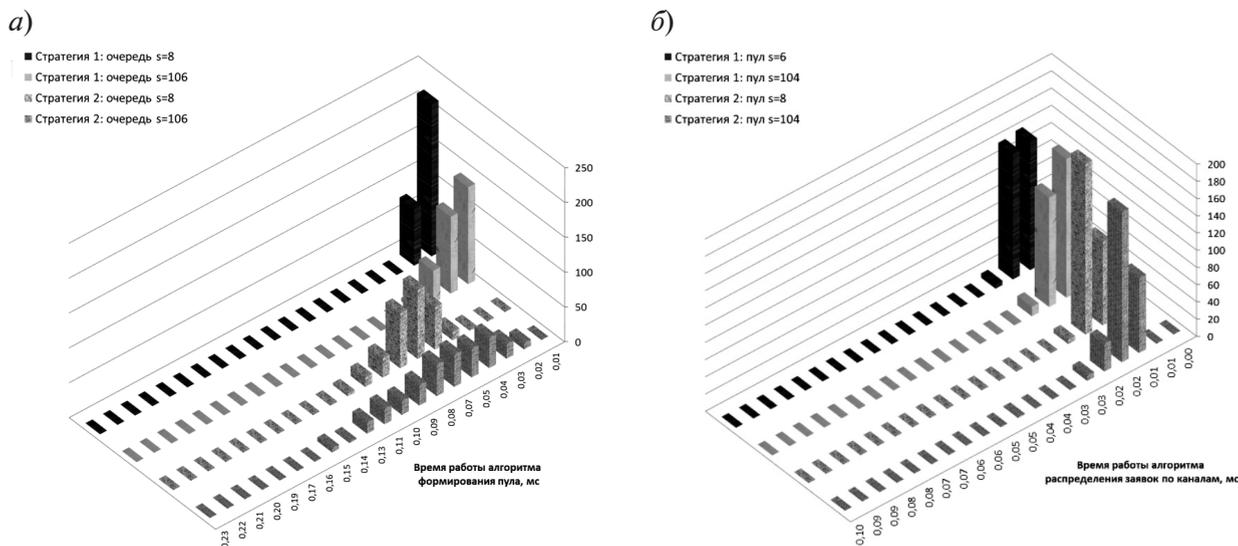


Рис. 7. Гистограммы распределения среднего времени обслуживания заявок на минимальных и максимальных значениях размера очереди (а) и пула (б)

На рис. 7 б представлены гистограммы распределения среднего времени обслуживания заявок на минимальных и максимальных значениях размера пула для алгоритма распределения заявок по каналам связи на 20 интервалах. Из гистограмм видно, что среднее время обслуживания заявок в пуле на максимальных и минимальных значениях не меняется для обеих стратегий, а сам алгоритм распределения заявок по каналам связи не зависит от выбранной стратегии формирования пула заявок.

Заключение

В статье представлен алгоритм, решающий двухкритериальную задачу поиска локального минимума между временем передачи потока данных по каналам передачи данных и пропускной способностью виртуального канала связи, удовлетворяющего минимальной требуемой пропускной способности заявки.

С целью подтверждения работоспособности и обоснования области применимости алгоритма динамического мультиплексирования гетерогенных каналов связи проведена серия экспериментов на определение зависимости среднего времени обслуживания заявок от размера очереди и пула. Для проведения экспериментов

выполнено компьютерное моделирование процесса распределения потоков данных между беспроводными гетерогенными сетями связи.

В качестве правил формирования пула были выбраны стратегии безусловного приоритета обслуживания очереди и «по квотам». Серия экспериментов показала, что среднее время обслуживания заявки:

не зависит от размеров очереди и пула, оно константно;

в очереди для алгоритма формирования пула для стратегии безусловного приоритета меньше, чем для стратегии «по квотам»;

в пуле для алгоритма распределения заявок по каналам передачи данных не зависит от выбранной стратегии формирования пула.

Развитием работы является расширение исследования других реализаций эвристик распределения потоков данных по каналам передачи данных и исследования времени исполнения алгоритма динамического мультиплексирования гетерогенных каналов связи на неполном наборе разрешенных каналов.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-29-04319.

СПИСОК ЛИТЕРАТУРЫ

1. **Hadama H., Chiba N., Fukui D., Ogura M., Nakamura R.** Implementation of dual transmission function for remote vehicle controls // Proc. of the 10th Asia-Pacific Symp. on Information and Telecommunication Technologies. Colombo, 2015. Pp. 1–3.
2. **Chirwa R.M.N., Lauf A.P.** Performance improvement of transmission in Unmanned Aerial Systems using multipath TCP // Proc. of the IEEE Internat. Symp. on Signal Processing and Information Technology. Noida, 2014. Pp. 000019–000024.
3. **Kumar Gupta S., Khara S.** Technique to improve the file transfer outcomes between road side unit and vehicles in vehicular ad-hoc networks // Proc. of the Internat. Conf. on Computing, Communication & Automation. Noida, 2015. Pp. 360–366.
4. **Sukuvaara T., Nurmi P.** Wireless traffic service platform for combined vehicle-to-vehicle and vehicle-to-infrastructure communications // IEEE Wireless Communications. 2009. Vol. 16. No. 6. Pp. 54–61.
5. **Viriyasitavat W., Bai F., Tonguz O.K.** Providing fairness between multiple flows in vehicular networks // Proc. of the Vehicular Networking Conf. IEEE. Kyoto, 2015. Pp. 44–51.
6. **Cordeschi N., Amendola D., Baccarelli E.** Reliable Adaptive Resource Management for Cognitive Cloud Vehicular Networks // IEEE Transactions on Vehicular Technology. 2015. Vol. 64. No. 6. Pp. 2528–2537.
7. **Popov S., Kurochkin M., Kurochkin L., Glazunov V.** The implementing of the internet of things concepts for the continuous provision of informational services for vehicle drivers and passengers // Proc. of the 1st Internat. Conf. on Telematics and Future Generation Networks. Pp. 1–5.
8. **Nigel Williams, Prashan Abeysekera, Nathan Dyer, Hai Vu** Multipath TCP in Vehicular to Infrastructure Communications // Grenville Armitage Centre for Advanced Internet Architectures. Technical Report 140828A. Swinburne University of Technology Melbourne.
9. **Ford A., Raiciu C., Handley M., Bonaventure O.** TCP Extensions for Multipath Operation with Multiple Addresses. RFC 6824. Internet Engineering Task Force // URL: <http://tools.ietf.org/html/rfc6824> (Дата обращения: Jan. 2013).
10. **Zhao L., Hong X., Gu B.** Handshake time and transmission rate of 802.11g measurement in vehicular networks // Proc. of the IEEE 35th Conf. on Local Computer Networks. 2010. Pp. 730–736.

Статья поступила в редакцию 27.05.2017

REFERENCES

1. **Hadama H., Chiba N., Fukui D., Ogura M., Nakamura R.** Implementation of dual transmission function for remote vehicle controls. *Proc. of the 10th Asia-Pacific Symposium on Information and Telecommunication Technologies*, Colombo, 2015, Pp. 1–3.
2. **Chirwa R.M.N., Lauf A.P.** Performance improvement of transmission in Unmanned Aerial Systems using multipath TCP. *Proc. of the IEEE International Symposium on Signal Processing and Information Technology*, Noida, 2014, Pp. 000019–000024.
3. **Kumar Gupta S., Khara S.** Technique to improve the file transfer outcomes between road side unit and vehicles in vehicular ad-hoc networks. *Proc. of the International Conference on Computing, Communication & Automation*, Noida, 2015, Pp. 360–366.
4. **Sukuvaara T., Nurmi P.** Wireless traffic service platform for combined vehicle-to-vehicle and vehicle-to-infrastructure communications. *IEEE Wireless Communications*, 2009, Vol. 16, No. 6, Pp. 54–61.
5. **Viriyasitavat W., Bai F., Tonguz O.K.** Providing fairness between multiple flows in vehicular networks. *Proc. of the Vehicular Networking Conference IEEE*, Kyoto, 2015, Pp. 44–51.
6. **Cordeschi N., Amendola D., Baccarelli E.** Reliable Adaptive Resource Management for Cognitive Cloud Vehicular Networks. *IEEE Transactions on Vehicular Technology*, 2015, Vol. 64, No. 6, Pp. 2528–2537.
7. **Popov S., Kurochkin M., Kurochkin L., Glazunov V.** The implementing of the internet of things concepts for the continuous provision of informational services for vehicle drivers and passengers. *Proc. of the 1st International Conference on Telematics and Future Generation Networks*, 2015, Pp. 1–5.
8. **Nigel Williams, Prashan Abeysekera, Nathan Dyer, Hai Vu** Multipath TCP in Vehicular to Infrastructure Communications. *Grenville Armitage Centre for Advanced Internet Architectures, Technical Report 140828A*, Swinburne University of Technology Melbourne.
9. **Ford A., Raiciu C., Handley M., Bonaventure O.** TCP Extensions for Multipath Operation with Multiple Addresses. *RFC 6824, Internet Engineering Task Force*. Available: <http://tools.ietf.org/html/rfc6824>

(Accessed: Jan. 2013).

10. **Zhao L., Hong X., Gu B.** Handshake time and transmission rate of 802.11g measurement

in vehicular networks. *Proc. of the IEEE 35th Conference on Local Computer Networks*, 2010, Pp. 730–736.

Received 27.05.2017

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

ПОПОВ Сергей Геннадьевич

POPOV Sergey G.

E-mail: popovserge@spbstu.ru

ТУЧКОВ Андрей Сергеевич

TUCHKOV Andrei S.

E-mail: tucha.te@gmail.com

КУРОЧКИН Леонид Михайлович

KUROCHKIN Leonid M.

E-mail: kurochkinl@spbstu.ru

ШАРАГИН Максим Павлович

SHARAGIN Maksim P.

E-mail: msharagin@gmail.com

DOI: 10.18721/JCSTCS.10302

УДК 621.391.28

АНАЛИЗ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ БЕСПРОВОДНЫХ ЦЕНТРАЛИЗОВАННЫХ СЕТЕЙ ДЛЯ ПЕРЕДАЧИ НЕАДАПТИВНОГО ВИДЕО ПО ПРОТОКОЛУ HTTP

И.А. Пастушок, А.М. Тюрликов

Санкт-Петербургский государственный университет аэрокосмического приборостроения,
Санкт-Петербург, Российская Федерация

Рост аудитории социальных сетей и их плотная интеграция с сервисами хранения и передачи видеоданных по протоколу HTTP приводит к колоссальному увеличению нагрузки на существующие телекоммуникационные системы. В связи с большой мобильностью современных пользователей настоящее явление наиболее актуально для беспроводных сетей связи, обладающих значительно меньшей производительностью по сравнению со стационарными аналогами. Большой объем видеоданных передается с использованием неадаптивной технологии передачи видео по протоколу HyperText Transfer Protocol (Facebook, трансляции в режиме реального времени и т. д.). Сочетание данных факторов приводит к необходимости исследования максимально возможной эффективности беспроводных сетей при передаче неадаптивного видеоконтента по протоколу HTTP. В статье рассмотрена модель беспроводной системы, позволяющая установить взаимосвязь между характеристиками сети и качеством воспроизведения неадаптивного видео при его передаче по протоколу HTTP. Описано нормированное отношение длительностей буферизации и просмотра, оказывающее критичное влияние на качество восприятия видеопотока. Для настоящего критерия найдена нижняя граница по всевозможным алгоритмам планирования распределения ресурсов беспроводного канала, характеризующая максимально возможную эффективность беспроводной системы при передаче неадаптивного видео по протоколу HTTP. На основе найденной границы предложен эвристический алгоритм распределения ресурсов беспроводного канала, реализующий концепцию совместного планирования, обладающий большей эффективностью по сравнению с существующими аналогами и близкой к найденной границе. Полученный результат может использоваться для формирования требований к алгоритмам планирования с учетом качества восприятия для существующих и последующих стандартов беспроводной связи.

Ключевые слова: передача видео по протоколу HTTP; DASH; беспроводные сети; эффективность; оптимизация; нижняя граница; планировщик.

Ссылка при цитировании: Пастушок И.А., Тюрликов А.М. Анализ эффективности использования беспроводных централизованных сетей для передачи неадаптивного видео по протоколу HTTP // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. Т. 10. № 3. 2017. С. 18–31. DOI: 10.18721/JCSTCS.10302

THE PERFORMANCE ANALYSIS OF CENTRALIZED WIRELESS NETWORKS FOR NON-ADAPTIVE HTTP-BASED VIDEO TRANSMISSION

I.A. Pastushok, A.M. Turlikov

St. Petersburg State University of Aerospace Instrumentation,
St. Petersburg, Russian Federation

At the present time, the rise in the social network audience and their tight integration with HTTP-based video streaming services leads to a sharp increase in the load on the current telecommunication networks. Due to high mobility of users, this effect has the greatest influence on wireless networks, which have rather low performance compared to hard-wired systems. The greatest amount of video traffic is transmitted using non-adaptive technology by HyperText Transfer Protocol (Facebook, a real-time event, etc.). This combination necessitates the investigation of the maximal possible performance of wireless networks for non-adaptive HTTP-based video streaming. In the study, we consider the wireless network model and propose the interrelation of its parameters and the streaming of non-adaptive HTTP-based video content. As the main performance criterion, we use the rebuffering percentage, which has a great influence on the user's Quality-of-Experience. For this criterion, we found the lower bound for all possible schedulers. In addition, based on the found lower bound, we proposed a heuristic scheduler that has a greater performance compared to the existing heuristic schedulers, close enough to the lower bound. The obtained result can be used as a baseline for the requirements to QoE-aware schedulers in the current and future wireless network generations.

Keywords: HTTP-based video; DASH; wireless networks; performance; optimization; lower bound; QoE-aware scheduler.

Citation: Pastushok I.A., Turlikov A.M. The Performance Analysis of Centralized Wireless Networks for Non-Adaptive HTTP-based Video Transmission. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems. 2017, Vol. 10, No. 3, Pp. 18–31. DOI: 10.18721/JCSTCS.10302

Введение

В настоящее время в телекоммуникационных сетях доминирует передача видеоданных. По оценкам ведущих компаний-производителей телекоммуникационного оборудования в 2016 г. видеотрафик составил 55 % от всего объема трафика, а к 2021 г. достигнет 60–75 % [1]. Подобное разделение объемов трафика приводит к специфичным требованиям к сетям передачи информации. Рост аудитории социальных сетей, плотно интегрированных с сервисами передачи и хранения неадаптивного видео по протоколу HyperText Transfer Protocol (HTTP), приводит к колоссальному росту нагрузки на уже существующие сети. В современных реалиях пользователи обладают высокой мобильностью, что приводит к актуальности рассмотрения беспроводных сетей связи в качестве носителей информации. В свою очередь беспроводные сети обладают значительно меньшей производительностью по сравнению со стационарными аналогами ввиду наличия беспроводного канала. Беспроводной канал характеризуется меньшими скоростями и высокими задержками при передаче информации. Более того, его состояние изменчиво во времени.

Важной задачей является оценка максимально возможной эффективности беспроводных сетей для передачи неадаптивных видеопотоков по протоколу HTTP. В работе [2] предложен критерий для оценки эффективности сетей передачи информации: нормированное отношение длительностей буферизации и просмотра (rebuffering percentage). Однако для данного критерия отсутствуют исследования максимально возможной эффективности беспроводных сетей при передаче видеоданных по протоколу HTTP.

Цель настоящей работы – нахождение количественной оценки максимально возможной эффективности беспроводных сетей связи для нормированного отношения длительностей буферизации и просмотра при передаче неадаптивного видео по протоколу HTTP.

Передача неадаптивных видеопоследовательностей в беспроводных сетях по протоколу HTTP

Последним этапом развития беспроводных систем связи являются беспроводные централизованные сети передачи информации. Под централизацией понимается на-

личие устройства (базовой станции), управляющего передачей данных абонентов, находящихся в зоне его ответственности. В настоящее время такие сети представлены в стандартах связи третьего и четвертого (LTE, LTE-A) поколений, также стандарты беспроводной связи пятого поколения будут централизованными.

В свою очередь передачу видео по протоколу HTTP регламентирует стандарт Dynamic Adaptive Streaming over HTTP (DASH) [3]. В качестве основных моментов стандарта DASH следует отметить его инвариантность к сети передачи информации. Модель системы передачи видеоданных по протоколу HTTP в беспроводных сетях связи представлена тремя основными компонентами (рис. 1):

- видео контент сервер – обеспечивает хранение и доступ к видеоданным;
- сеть передачи информации (магистральная сеть, базовая станция, радиоканал) – телекоммуникационная сеть, построенная на произвольном физическом принципе, обеспечивающая корректную работу протокола HTTP;
- воспроизводящие устройства – программно-аппаратные комплексы, обеспечивающие загрузку и демонстрацию видео с контент сервера. В модели каждое устройство характеризуется уникальным номером $i = \overline{1, N}$, где N – общее число воспроизводящих устройств в системе.

Стандарт DASH описывает две технологии организации передачи видео: адаптивную и неадаптивную. В рамках настоящей статьи описано исследование неадаптивной технологии (HTTP Progressive Download), т. к. большие объемы видеотрафика передаются с использованием данной технологии (Facebook, трансляции в режиме реального

времени и т. д.).

Далее в статье производится формализация компонентов модели путем введения основных понятий и допущений.

Формат представления видеоданных. При передаче видео по протоколу HTTP все видеопоследовательности представляются набором сегментов одинаковой длительности, равной d секунд. Длительность видео является независимой случайной величиной с конечным математическим ожиданием и дисперсией. Каждый сегмент, запрашиваемый пользователем i , характеризуется порядковым номером j и представлен в дискретном наборе битовых скоростей $\forall i, j : R_{i,j} \in \{R_{\min}, \dots, R_{\max}\}$.

При передаче сегмента j пользователем i он представляется последовательностью из P_{ij} пакетов равного объема, что обусловлено работой протокола транспортного уровня.

Модель поведения пользователя. При рассмотрении передачи неадаптивного видео каждый пользователь в начальный момент времени выбирает битовую скорость потока R_i , качество которой его абсолютно устраивает, и запрашивает все сегменты видео с выбранной битовой скоростью: $\forall i : R_{i,1} = R_{i,2} = \dots = R_i$.

Пользователь просматривает последовательность из роликов. В начале каждого ролика воспроизводящим устройством производится начальная буферизация, после чего начинается непосредственно демонстрация видео. Если скорость получения информации по каналу связи будет меньше битовой скорости запрашиваемого видео, то пользователь будет наблюдать прерывание воспроизведения потока, вызванное повторными буферизациями. После окончания просмотра ролика каждый пользова-

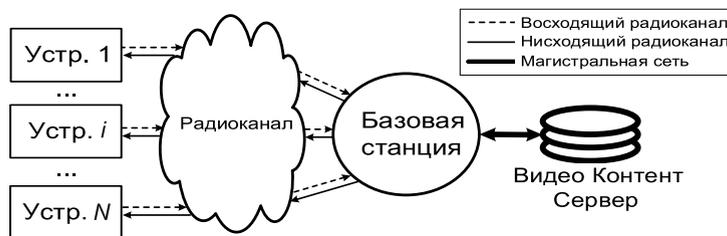


Рис. 1. Беспроводная централизованная система передачи видеоданных по протоколу HTTP

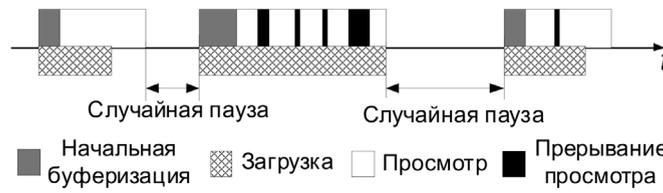


Рис. 2. Модель поведения пользователя при просмотре видео

тель i ждет случайную паузу перед заказом следующего, которая является независимой случайной величиной с конечными математическим ожиданием и дисперсией (рис. 2). В настоящей модели пользователь считается *активным* в момент времени t , если в данный момент он осуществляет загрузку данных, в противном случае он считается *неактивным*.

Подобную модель поведения для каждого пользователя i возможно охарактеризовать двумя величинами:

Коэффициент разреженности видеопотока – отношение суммы длительностей просмотра (w_i^T) и пауз между просмотрами (p_i^T) к длительности просмотра за длительное время T :

$$\gamma_i = \lim_{T \rightarrow \infty} \frac{w_i^T + p_i^T}{w_i^T}.$$

Нормированное отношение длительностей буферизации и просмотра – отношение общей длительности буферизации b_i^T (начальные буферизации и прерывания просмотра) к сумме длительностей буферизации и просмотра за продолжительное время T :

$$g_i = \lim_{T \rightarrow \infty} \frac{b_i^T}{b_i^T + w_i^T}.$$

В реальных системах при просмотре видео существует зависимость между длительностями просмотра и паузами между ними: после просмотра длительных и коротких роликов пользователь осуществляет длительную и короткую паузу соответственно. Следствием настоящего факта является равенство коэффициентов разреженности потоков для всех пользователей: $\forall i: \gamma_i = \gamma$.

Модель беспроводного канала. В данной статье рассматриваются беспроводные слотовые системы: все время работы подобных систем разделено на слоты рав-

ной длительности. В каждом слоте имеется ограниченное число частотно-временных ресурсов, зависящее от ширины доступной полосы передачи. В качестве исследуемой модели беспроводного канала связи рассматривается канал с плоскими замираниями: в каждый момент времени затухание распространения сигнала для каждого конкретного пользователя не зависит от частоты передачи, следовательно, в каждом частотно-временном ресурсе пользователь может передать одинаковый объем данных. В рамках настоящей работы предполагается, что настройки модуляции и помехоустойчивого кодирования подбираются так, что вероятность ошибки или потери пакета при передаче пренебрежимо мала. Таким образом, для каждого пользователя i в момент времени t состояние беспроводного канала связи возможно охарактеризовать *максимально достижимой скоростью передачи данных в канале* $C_i(t)$ – скорость передачи информации в беспроводном канале при условии, что все ресурсы канала в момент времени t были выделены пользователю i .

Предполагается, что состояние беспроводного канала изменяется таким образом, что в течение загрузки пакета k из сегмента j максимальная достижимая скорость канала пользователя i постоянна:

$$C_i(t) \Big|_{t_{i,j,k}}^{t_{i,j,k} + \Delta t_{i,j,k}} = C_{i,j,k},$$

где $t_{i,j,k}$ – момент времени начала загрузки пользователем i пакета k из сегмента j ; $\Delta t_{i,j,k}$ – длительность загрузки пользователем i пакета k из сегмента j ; $C_{i,j,k}$ – максимальная достижимая скорость канала пользователя i в течение загрузки пакета k из сегмента j .

Введем дополнительное обозначение $C_{i,j}^{-1}$:

$$C_{i,j}^{-1} = \frac{1}{P_{i,j}} \sum_{k=1}^{P_{i,j}} \frac{1}{C_{i,j,k}}, \quad (1)$$

где $P_{i,j}$ – количество пакетов в сегменте j пользователя i .

На последовательность случайных величин $C_{i,1}^{-1}, C_{i,2}^{-1}, \dots, i = \overline{1, N}$ накладывается ограничение, что она формирует эргодический случайный процесс с конечным математическим ожиданием $E[C_i^{-1}]$.

Модель планировщика ресурсов беспроводного канала. Определяющий вклад в эффективность беспроводной системы в целом вносит алгоритм распределения ресурсов радиоканала на базовой станции, т. к. скорость передачи информации определяется скоростью передачи в самом узком месте, которым в настоящей системе является радиоканал. От того, каким образом будут распределяться ресурсы канала, будут зависеть скорости загрузки информации пользовательскими устройствами, удовлетворенность пользователя обслуживанием и, как следствие, эффективность системы в целом. Решением задачи распределения ресурсов беспроводного канала связи занимается алгоритм планирования распределения ресурсов на канальном уровне, называемый *планировщиком*.

При рассмотрении беспроводного канала с плоскими замираниями работу алгоритма планирования в каждый момент времени t возможно описать набором долей канала, выделенным пользователям:

$$A(t) = \{\alpha_i(t), i = \overline{1, N}\}.$$

Очевидным ограничением на возможные значения функций $\alpha_i(t)$ является следующее неравенство:

$$\forall t : \sum_{i=1}^N \alpha_i(t) \leq 1. \quad (2)$$

Неравенство (2) может быть интерпретировано следующим образом: в любой момент времени работы алгоритма планирования общий объем выделенных ресурсов не превышает доступного объема ресурсов для канала передачи информации. Следовательно, для любого пользователя i мгновенная

скорость передачи информации в нисходящем канале связи $S_i(t)$ может быть вычислена следующим образом:

$$S_i(t) = \alpha_i(t)C_i(t). \quad (3)$$

Алгоритм планирования в каждый момент времени решает задачу распределения ресурсов беспроводного канала. Для решения данной задачи ему доступна информация о предыстории, а именно доли выделенных ресурсов канала, значения максимально достижимой скорости и объем переданной информации для каждого пользователя:

$$A(t) = \mathcal{A}(S_i(\tau), C_i(\tau); \tau < t, i = \overline{1, N}), \quad (4)$$

где $\mathcal{A}(\cdot)$ является алгоритмом планирования.

В формуле (4) информация о предыстории выделенных долей беспроводного канала и объемах переданной информации для пользователя i агрегированы в значении $S_i(\tau)$, так как данные параметры задаются соотношением (3). На работу алгоритма планирования накладывается ряд допущений:

- алгоритм планирования выделяет частотно-временные ресурсы только активным абонентам;
- обеспечивает каждому активному абоненту минимальную долю ресурсов беспроводного канала, отличную от нуля в каждом интервале планирования;
- распределяет все доступные частотно-временные ресурсы.

Ввиду большого числа используемых обозначений в таблице агрегированы важные обозначения с расшифровками и единицами измерения.

Взаимосвязь между характеристиками системы передачи информации и воспроизведением видеоряда

В [4] представлена аналогичная модель системы передачи видеоданных в беспроводных системах и предложена взаимосвязь между характеристиками системы передачи информации и воспроизведением видеоряда, однако, в данной работе рассматривался канал связи, характеристики которого неизменны во времени. В настоящей статье

Используемые обозначения

Обозначение	Описание	Единица измерения
N	Число пользователей в системе	шт.
индекс i	Порядковый номер пользователя	шт.
индекс j	Порядковый номер сегмента видеоданных	шт.
индекс k	Порядковый номер пакета в сегменте	шт.
d	Длительность сегмента видеоданных	с
R_i	Битовая скорость, выбранная пользователем i	бит/с
P_{ij}	Число пакетов в сегменте j пользователя i	шт.
b_i^T	Длительность буферизаций пользователя i в течение времени T	с
w_i^T	Длительность просмотра пользователя i в течение времени T	с
p_i^T	Длительность пауз между просмотрами роликов пользователя i в течение времени T	с
g_i	Нормированное отношение длительностей буферизации и просмотра пользователя i	Безразмерная величина
γ	Коэффициент разреженности видеопотока	Безразмерная величина
C	Максимально достижимая скорость передачи данных в канале	бит/с
α	Выделенная доля ресурсов беспроводного канала	Безразмерная величина
S	Скорость передачи данных по беспроводному каналу	бит/с

предлагается обобщение модели и взаимосвязи, представленных в [4] для канала, характеристики которого изменяются во времени.

Утверждение. Для всевозможных алгоритмов планирования, удовлетворяющих системе допущений, следующее неравенство истинно:

$$\sum_{i=1}^N R_i E[C_i^{-1}] \frac{1 - g_i}{g_i + \gamma(1 - g_i)} \leq 1.$$

Доказательство. Рассмотрим передачу одного пакета k из сегмента под номером j пользователем i . При загрузке настоящего пакета через беспроводной канал было передано $\frac{R_i d}{P_{i,j}}$ бит видеоданных. Эта величина может быть вычислена на основе наблюдения за скоростью передачи информации в беспроводном канале:

$$\int_{t_{i,j,k}}^{t_{i,j,k} + \Delta t_{i,j,k}} S_i(t) dt = C_{i,j,k} \int_{t_{i,j,k}}^{t_{i,j,k} + \Delta t_{i,j,k}} \alpha_i(t) dt = \frac{R_i d}{P_{i,j}}.$$

Перенесем $C_{i,j,k}$ в правую часть последнего равенства и получим долю времени использования канала пользователем i при загрузке k -го пакета:

$$\int_{t_{i,j,k}}^{t_{i,j,k} + \Delta t_{i,j,k}} \alpha_i(t) dt = \frac{R_i d}{P_{i,j}} \frac{1}{C_{i,j,k}}.$$

Просуммировав левую и правую части по всем пакетам в сегменте, получим долю времени использования канала пользователем i при загрузке j -го сегмента:

$$\int_{t_{i,j}}^{t_{i,j} + \Delta t_{i,j}} \alpha_i(t) dt = R_i d \left(\frac{1}{P_{i,j}} \sum_{k=1}^{P_{i,j}} \frac{1}{C_{i,j,k}} \right),$$

где $t_{i,j}$ — время начала загрузки пользователем

лем i сегмента j ; $\Delta t_{i,j}$ – длительность загрузки пользователем i сегмента j .

Используя равенство (1), получим следующее выражение:

$$\int_{t_{i,j}}^{t_{i,j}+\Delta t_{i,j}} \alpha_i(t) dt = R_i C_{i,j}^{-1} d.$$

Тогда доля времени использования канала пользователем i за время T :

$$\int_0^T \alpha_i(t) dt = d \sum_{j=1}^{H_i^T} R_i C_{i,j}^{-1} + \Delta s_i, \quad (5)$$

где H_i^T – число полностью загруженных сегментов пользователем i в течение времени T ; Δs_i – доля времени использования канала пользователем i , затраченная на загрузку невоспроизведенного сегмента видео.

Просуммировав уравнение (5) для всех пользователей в системе ($i = \overline{1, N}$) и внося сумму под знак интеграла, получим следующее выражение:

$$\int_0^T \sum_{i=1}^N \alpha_i(t) dt = \sum_{i=1}^N d \sum_{j=1}^{H_i^T} R_i C_{i,j}^{-1} + \Delta s,$$

где Δs – доля времени использования канала всеми пользователями, затраченная на загрузку невоспроизведенных сегментов видео.

Так как в каждый момент времени t сумма выделенных долей канала пользователям не может превышать единицы, согласно условию (2), следовательно

$$\int_0^T \sum_{i=1}^N \alpha_i(t) dt \leq T :$$

$$d \sum_{i=1}^N \sum_{j=1}^{H_i^T} R_i C_{i,j}^{-1} + \Delta s \leq T.$$

Разделим левую и правую части неравенства на T и внесем его под первый знак суммы. Так как $T > 0$, знак неравенства не изменится:

$$\sum_{i=1}^N \frac{d \sum_{j=1}^{H_i^T} R_i C_{i,j}^{-1} + \Delta s}{T} \leq 1.$$

Для каждого пользователя время пребывания в системе состоит из буферизаций, просмотров и пауз между просмотрами ви-

део $T = b_i^T + w_i^T + p_i^T$. Следовательно, последнее неравенство представляется в следующем виде:

$$\sum_{i=1}^N \frac{d \sum_{j=1}^{H_i^T} R_i C_{i,j}^{-1} + \Delta s}{b_i^T + w_i^T + p_i^T} \leq 1.$$

Так как длительность просмотра равняется длительности всех загруженных сегментов, $w_i^T = H_i^T d$, $i = \overline{1, N}$:

$$\sum_{i=1}^N \frac{w_i^T}{H_i^T} \frac{1}{b_i^T + w_i^T + p_i^T} \sum_{j=1}^{H_i^T} R_i C_{i,j}^{-1} + \Delta s \leq 1.$$

Разделим числитель и знаменатель левой части полученного выражения на w_i^T и устремим время в бесконечность ($T \rightarrow \infty$). Для каждого абонента $i = \overline{1, N}$ воспользуемся определением g_i , свойствами эргодичности процессов $C_{i,j}^{-1}$, а именно существованием

предела $\lim_{H_i^T \rightarrow \infty} \left[(H_i^T)^{-1} \sum_{j=1}^{H_i^T} C_{i,j}^{-1} \right] = E[C_i^{-1}]$, и получим следующее выражение:

$$\sum_{i=1}^N R_i E[C_i^{-1}] \frac{1 - g_i}{g_i + \gamma(1 - g_i)} \leq 1.$$

Настоящий факт завершает доказательство.

Нижняя граница для нормированного отношения длительностей буферизации и просмотра при передаче неадаптивного видео

Важным аспектом настоящего исследования является выбор критерия эффективности системы. В современных реалиях эффективность телекоммуникационных сетей связана с удовлетворенностью пользователя, поэтому в данной статье будет проведен краткий анализ факторов, влияющих на удовлетворенность пользователя при просмотре видео по протоколу HTTP.

В работах [5, 6] выделено два основных фактора, влияющих на качество восприятия видео:

- битовая скорость потока, определяющая разрешение воспроизводимого видео [7–9];
- длительность буферизации, когда

пользователь ожидает начала или возобновления просмотра [2, 4, 10–13].

Так как в настоящей статье предполагается, что пользователь выбирает качество видеопотока исходя из собственных предпочтений, то на его удовлетворенность может оказывать влияние только длительность буферизации при просмотре видео. На основе проведенного обзора работ в настоящей предметной области сделан вывод о том, что удовлетворенность пользователя обратно пропорциональна длительности буферизаций.

В рамках этой работы за критерий качества восприятия для одного пользователя принято нормированное отношение длительностей буферизации и просмотра, предложенное в работе [2], а в качестве критерия эффективности системы в целом рассматривается среднее значение выбранного критерия по всему множеству пользователей.

На данную величину непосредственное влияние оказывает алгоритм планирования A , установленный на базовой станции, т. к. он определяет скорости передачи информации. Для каждого конкретного алгоритма планирования A возможно вычислить среднее значение нормированного отношения длительностей буферизации и просмотра:

$$\bar{g}(A) = \frac{1}{N} \left(\sum_{i=1}^N g_i(A) \right).$$

Так как удовлетворенность пользователей в системе обратно пропорциональна значению $\bar{g}(A)$, то в настоящей работе будет найдена нижняя граница по всевозможным алгоритмам планирования A , удовлетворяющим системе допущений, характеризующая максимально возможную эффективность системы по введенному критерию:

$$G = \inf_{A \in \mathcal{A}} \bar{g}(A). \quad (6)$$

Задача нахождения нижней границы по всевозможным алгоритмам планирования сводится к нахождению решения оптимизационной задачи, основанной на утверждении и выражении (6), решение которой

характеризует максимально возможную эффективность беспроводной централизованной сети при передаче неадаптивного видео по протоколу НТТР:

$$\text{Минимизировать: } G = \sum_{i=1}^N g_i$$

При условии:

$$\begin{cases} \sum_{i=1}^N \frac{K_i(1-g_i)}{g_i + \gamma(1-g_i)} \leq 1, \\ g_i \in [0, 1], i = \overline{1, N} \end{cases} \quad (7)$$

где $K_i = R_i E[C_i^{-1}]$.

Оптимизационная задача является задачей нелинейного выпуклого программирования и может быть сведена к обобщенной задаче о непрерывном рюкзаке [14], в которой в качестве целевой функции выступает сумма монотонно возрастающих выпуклых функций:

$$\text{Максимизировать: } \sum_{i=1}^N f(x_i)$$

При условии:

$$\begin{cases} \sum_{i=1}^N w_i x_i \leq W \\ W \in (0, \infty) \\ x_i \in [0, 1], \quad i = \overline{1, N} \\ w_i \in (0, \infty), \quad i = \overline{1, N} \end{cases} \quad (8)$$

где $f(x_i)$ – непрерывная, дважды дифференцируемая, строго монотонно возрастающая, выпуклая вниз функция, ограниченная сверху и снизу в области определения x_i .

Решение задачи (8) может быть получено с помощью алгоритма 1, обладающего низкой вычислительной сложностью ($O(N \log_2 N)$, $N \rightarrow \infty$) по сравнению со стандартными методами решения [15]. Данный факт является следствием того, что задача (8) обладает свойством жадного выбора, аналогично задаче о непрерывном рюкзаке [14].

Алгоритм 1: Решение задачи (8)

1. Сортировка предметов по возрастанию значений w_i и их нумерация в соответствии с сортировкой.
2. Нахождение предмета с индексом

$$j : \begin{cases} \sum_{i=1}^{j-1} w_i \leq W \\ \sum_{i=1}^j w_i > W \end{cases}$$

если $j \in \emptyset, j = N + 1$.

3. Нахождение оптимальных значений:

$$x_i = \begin{cases} 1, & i < j \\ \frac{1}{w_j} \left(W - \sum_{i=1}^{j-1} w_i \right), & i = j \\ 0, & i > j \end{cases}$$

В рамках настоящей работы задача (7) будет сведена к задаче (8) и предложен алгоритм нахождения нижней границы для критерия G.

Теорема. Решение задачи нелинейного программирования (7), характеризующее нижнюю границу нормированного отношения длительностей буферизации и просмотра при передаче неадаптивных видеопотоков, может быть получено с помощью алгоритма 2.

Алгоритм 2: Решение задачи (7)

1. Сортировка пользователей по возрастанию отношений $\left(\frac{K_i}{\gamma}\right)$ и их нумерация в соответствии с сортировкой.

2. Нахождение пользователя с индексом

$$j : \begin{cases} \sum_{i=1}^{j-1} \frac{K_i}{\gamma} \leq 1 \\ \sum_{i=1}^j \frac{K_i}{\gamma} > 1 \end{cases}$$

если $j \in \emptyset, j = N + 1$.

3. Нахождение оптимальных значений g_i :

$$g_i = \begin{cases} 0, & i < j \\ \frac{\gamma(1-\xi)}{\xi + \gamma(1-\xi)}, & i = j \\ 1, & i > j \end{cases}$$

где $\xi = \frac{\gamma}{K_j} \left(1 - \sum_{i=1}^{j-1} \frac{K_i}{\gamma} \right)$.

Доказательство. Произведем несколько преобразований системы (7). Изначально вынесем из знаменателя нелинейного ограничения γ :

$$\sum_{i=1}^N \frac{K_i(1-g_i)}{g_i + \gamma(1-g_i)} = \sum_{i=1}^N \left(\frac{K_i}{\gamma} \right) \frac{(1-g_i)}{\frac{g_i}{\gamma} + (1-g_i)}.$$

Далее произведем преобразование задачи (7) из задачи минимизации в задачу максимизации:

Максимизировать: $\sum_{i=1}^N (1-g_i)$

При условии:

$$\begin{cases} \sum_{i=1}^N \left(\frac{K_i}{\gamma} \right) \frac{(1-g_i)}{\frac{g_i}{\gamma} + (1-g_i)} \leq 1 \\ g_i \in [0, 1], i = \overline{1, N} \end{cases}$$

Введем обозначение $x_i = (1-g_i) / \left(\frac{g_i}{\gamma} + (1-g_i) \right)$. Значение x_i монотонно зависит от величины g_i и может принимать значения в отрезке $[0, 1]$. Выразим значения g_i и $(1-g_i)$ из x_i :

$$g_i = \frac{\gamma(1-x_i)}{x_i + \gamma(1-x_i)}; \quad 1-g_i = \frac{x_i}{x_i + \gamma(1-x_i)}.$$

Основываясь на введенных обозначениях, оптимизационная задача (7) принимает следующий вид:

Максимизировать: $\sum_{i=1}^N \frac{x_i}{x_i + \gamma(1-x_i)}$

При условии:

$$\begin{cases} \sum_{i=1}^N \left(\frac{K_i}{\gamma} \right) x_i \leq 1 \\ x_i \in [0, 1], i = \overline{1, N} \end{cases} \quad (9)$$

Легко показать, что целевая функция задачи (9) соответствует требованиям к целевой функции задачи (8). Основываясь на настоящем факте, решение задачи (9) может быть получено при использовании алгоритма 1, где $w_i = \left(\frac{K_i}{\gamma}\right), i = \overline{1, N}$ и $W = 1$. Таким образом, задача (7) сведена к задаче (8), а алгоритм 2 является адаптацией алгоритма 1 в обозначениях оптимизационной задачи (7) и взаимосвязи между значениями g_i и x_i . Данный факт завершает доказательство.

Обобщая сказанное ранее, в результате была найдена величина G , характеризующая максимально возможную эффективность алгоритмов распределения ресурсов радиоканала в централизованных беспроводных телекоммуникационных сетях при передаче неадаптивных видеопотоков.

Эвристический алгоритм планирования для минимизации нормированного отношения длительностей буферизации и просмотра

В настоящей статье предлагается концепция построения алгоритмов планирования распределения ресурсов беспроводного канала с учетом требований к качеству восприятия абонента: *совместное планирование*. В классических планировщиках: Round Robin, обеспечивающий равный доступ к ресурсам канала связи, и Proportional Fair, обеспечивающий равные скорости передачи информации пользователям, приоритет на частотно-временные ресурсы для каждого пользователя вычисляется незави-

симо, и далее пользователь с максимальным приоритетом получает данный ресурс. В рамках этого исследования предлагается разработать эвристический планировщик, который при вычислении приоритета будет учитывать характеристики беспроводного канала и загружаемого видеоряда всего множества абонентов в соте. Данная концепция позволяет увеличить эффективность алгоритмов планирования для передачи неадаптивных видеопотоков по протоколу HTTP.

В предлагаемом планировщике будет использован алгоритм 2, позволяющий произвести вычисление оценки нижней границы для критерия G на основе оценок параметров сети передачи информации и просматриваемых пользователями видеоданных. В результате работы алгоритма 2 будут доступны значения $\hat{g}_i, i = \overline{1, N}$, определяющие оценку минимально возможного значения критерия G . Используя данные значения, возможно вычислить рекомендуемые скорости передачи информации для каждого пользователя:

$$S_i^{rec}(t) = \begin{cases} \hat{R}_i(t), & \hat{g}_i(t) = 0 \\ \hat{R}_i(t)(1 - \hat{g}_i(t)), & \hat{g}_i(t) \in (0, 1), i = \overline{1, N}, \\ S^{\min}, & \hat{g}_i(t) = 1 \end{cases}$$

где S^{\min} – минимальная скорость передачи информации, обеспечиваемая алгоритмом планирования, обусловленная требованиями стандартов связи; $\hat{R}_i(t)$ – оценка битовой скорости просматриваемого видеопотока пользователем i в момент времени t , доступной от системы анализа трафика, встроенной в опорную сеть оператора.

Таким образом, появляется задача обеспечения рекомендуемых скоростей получения информации на базовой станции. Для решения настоящей задачи предлагается в каждом слоте оценивать две характеристики для каждого пользователя: максимально достижимую скорость канала и скорость передачи информации в каждый момент времени – $C_i^{avg}(t)$ и $S_i^{avg}(t)$ соответственно по следующим методикам:

$$C_i^{avg}(t) = \left(1 - \frac{1}{w_C(t)}\right) C_i^{avg}(t-1) +$$

$$+ \left(\frac{1}{w_C(t)}\right) C_i(t).$$

$$S_i^{avg}(t) = \left(1 - \frac{1}{w_S(t)}\right) S_i^{avg}(t-1) + \left(\frac{1}{w_S(t)}\right) \frac{P_i(t)}{\Delta t},$$

где $w_C(t) = \min(t - t_i^l + 1, \bar{w}_C)$, $w_S(t) = \min(t - t_i^l + 1, \bar{w}_S)$; t_i^l – последний момент времени, когда для пользователя i была информация для передачи; \bar{w}_C и \bar{w}_S – длительности интервалов усреднения для оценок C_i^{avg} и S_i^{avg} соответственно; $P_i(t)$ – количество бит, которое пользователь i передал в момент времени t ; Δt – длительность слота; $S_i^{avg}(0) = C_i^{avg}(0) = 0$.

Данные методики основаны на оценке среднего значения методом экспоненциального забывания с изменяемой длитель-

ностью усреднения. Так как пользователь делает паузы между просмотрами, то будем считать пользователя активным в момент времени t , если у него есть данные для передачи на базовой станции, иначе будем считать пользователя неактивным.

Для обеспечения рекомендуемой скорости передачи информации модифицируем метод вычисления приоритета планировщика Proportional Fair:

$$p_{r,i}(t) = 1\{S_i^{avg}(t) \leq S_i^{rec}(t)\} \frac{C_i^{avg}(t)}{S_i^{avg}(t)}, \quad (10)$$

где $p_{r,i}(t)$ – приоритет пользователя i на частотно-временной ресурс r в момент времени t , $1\{\}$ – индикаторная функция.

Подобная модификация не позволяет алгоритму планирования выделять ресурсы пользователю, чья оценка скорости передачи превышает рекомендуемую скорость получения информации. Такой подход позволяет обеспечить скорость получения информации пользователем, равную рекомендуемой.

На основе описанного выше предлагается алгоритм 3, который минимизирует значение критерия G .

Алгоритм 3. Совместное планирование распределения ресурсов радиоканала для минимизации критерия G в каждом слоте.

1. Определить список активных пользователей, у которых имеются данные для передачи на базовой станции.

2. Вычислить оценки нижней границы $\hat{g}_i, i = \overline{1, N}$ при $K_i = \hat{R}_i(t) / C_i^{avg}(t)$ и $\gamma = 1$ (алгоритм 2).

3. Для каждого активного пользователя вычислить рекомендуемые скорости передачи информации, используя вычисленные значения оценки нижней границы на Шаге 2.

4. Для каждого частотно-временного ресурса найти активного пользователя с максимальным приоритетом в соответствии с выражением (10).

5. Выделить пользователям с максимальным приоритетом частотно-временные ресурсы.

6. Обновить значения C_i^{avg} и S_i^{avg} для $i = \overline{1, N}$.

Численный пример. Для демонстрации эффективности предложенного планировщика в сравнении с существующими эвристическими и нижней границей было проведено моделирование в низкоуровневой системе моделирования NS-3, полностью соответствующей стандарту LTE. В данной системе моделирования был имплементирован комплекс оконечных устройств, реализующий передачу видеоданных по протоколу DASH [3, 16].

Пользователи расставлены на линии по удалению от базовой станции с равным расстоянием между соседними устройствами с максимальным удалением от базовой станции равным 700 м. Затухание при распространении сигнала описывается моделью городской застройки Окамура-Хата. Плоские замирания канала обусловлены движением окружения со скоростью 3 км/ч. Все пользователи смотрят ролики длительностью 300 с с битовой скоростью 1 Мбит/с. Паузы между роликами распределены по усеченному экспоненциальному закону с математическим ожиданием 30 с в отрезке [15,45] с. Алгоритм планирования гарантирует каждому пользователю скорость получения информации не менее 100 Кбит/с.

В представленном сценарии увеличивалось число пользователей и анализировалась эффективность алгоритмов планирования (рис. 3). Из результатов, представленных на рисунке, следует, что предложенный алгоритм планирования обеспечивает меньшие значения критерия G по сравнению с существующими аналогами и значения, близкие к найденной нижней границе. Также при зафиксированном значении критерия качества предложенный планировщик обеспечивает на 7–14 % больше удовлетворенных пользователей относительно известных эвристик и, как следствие, обладает большей эффективностью.

Заключение

В статье рассмотрена модель передачи неадаптивного видео по протоколу HTTP в беспроводных централизованных системах, в которой условия беспроводного канала связи изменяются во времени. Для дан-

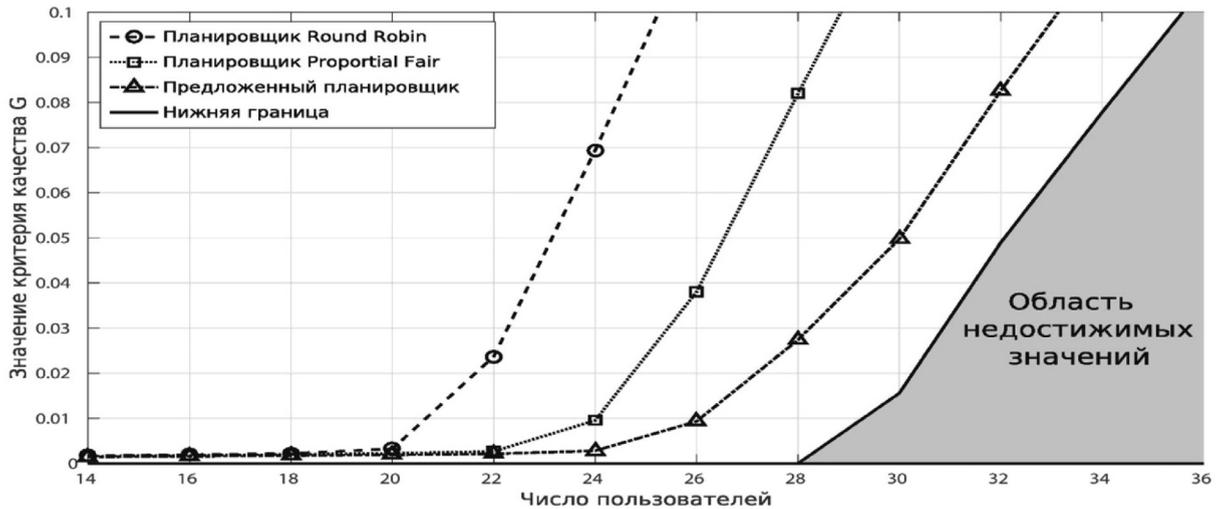


Рис. 3. Сравнение эффективности алгоритмов планирования с найденной нижней границей

ной модели предложена взаимосвязь между характеристиками системы передачи информации и воспроизведением видеоряда. Основываясь на найденной взаимосвязи, найдена нижняя граница по всевозможным алгоритмам планирования распределения ресурсов беспроводного канала, удовлетворяющим введенной модели, для нормированного отношения длительностей буферизации и просмотра.

Предложен эвристический алгоритм планирования, обладающий эффективностью на 7–14 % больше, чем известные эвристики. Эффективность предложенного

планировщика близка к найденной нижней границе.

Полученный результат позволяет увеличить эффективность беспроводных систем для передачи видеоданных по протоколу HTTP и сформировать требования к алгоритмам планирования, учитывающие качество восприятия контента, для текущего и последующих поколений беспроводных сетей.

Работа выполнена в рамках инициативного научного проекта № 8.8540.2017/БЧ «Разработка алгоритмов передачи данных в системах IoT с учетом ограничений на сложность устройств».

СПИСОК ЛИТЕРАТУРЫ

1. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2016–2021 White Paper. 2017 // URL: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html> (Дата обращения: 01.04.2017).
2. Oyman O., Singh S. Quality of Experience for HTTP Adaptive Streaming Services // IEEE Communication Magazine. 2012. No. 4. Pp. 20–27. DOI: 10.1109/issn0163-6804
3. 3GPP TS 26.247 v14.0.0. Transparent End-to-End Packet Switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming Over HTTP (3GP-DASH) // 3GPP. 2016. 104 p.
4. Bakin E., Evseev G., Turlikov A. Lower Bound of Mean user Waiting/Watching Time Ratio in Wireless Video Streaming Network // IEEE Global Communications Conf. Austin TX, 2014. Pp. 1170–1175.
5. Video as a Basic Service of LTE Networks: Mobile vMOS Defining Network Requirements // URL: <http://www.huawei.com/minisite/4-5g/en/industryjsdc-j.html> (Дата обращения: 01.04.2017).
6. P.1201. Amendment 2, Parametric Non-intrusive Assessment of Audiovisual Media Streaming Quality Amendment 2: New Appendix III Use of ITU-T P.1201 for Non-adaptive, Progressive Download Type Media Streaming. ITU-T, 2013. 23 p.
7. Essaili A.E., Schroeder D., Staehle D. et al. Quality-of-Experience Driven Adaptive HTTP Media Delivery // IEEE ICC 2013 – Communication QoS, Reliability and Modeling Symp. Budapest, 2013. Pp. 2480–2485.
8. Pu W., Zou Z., Chen C.W. Video Adaptation

Proxy for Wireless Dynamic Adaptive Streaming over HTTP // Proc. of 19th Intern. Packet Video Workshop. Munich, 2012. Pp. 65–70.

9. **Chen Y.C., Chang J.W., Wei H.Y.** A Multi-level QoE Framework for Smartphone Video Streaming Applications // 2014 Globecom Workshops. Austin TX, 2014. Pp. 225–230.

10. **Bui N., Valentin S., Widmer J.** Anticipatory Quality-Resource Allocation for Multi-user Mobile Video Streaming // IEEE Conf. on Computer Communications Workshops. Hong Kong, 2015. Pp. 245–250.

11. **Khan A., Sun L., Jammeh E., Ifeakor E.** Quality of Experience-Driven Adaptation Scheme for Video Applications over Wireless Networks // IET Communications. 2010. No. 11. Pp. 1337–1347. DOI: 10.1049

Статья поступила в редакцию 03.07.2017

REFERENCES

1. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2016–2021 White Paper. 2017.* Available: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html> (Accessed: 01.04.2017).

2. **Oyman O., Singh S.** Quality of Experience for HTTP Adaptive Streaming Services. *IEEE Communication Magazine*, 2012, No. 4, Pp. 20–27. DOI: 10.1109/issn0163-6804

3. 3GPP TS 26.247 v14.0.0. Transparent End-to-End Packet Switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming Over HTTP (3GP-DASH). *3GPP*, 2016, 104 p.

4. **Bakin E., Evseev G., Turlikov A.** Lower Bound of Mean user Waiting/Watching Time Ratio in Wireless Video Streaming Network. *Proc. of the IEEE Global Communications Conf.*, Austin TX, 2014, Pp. 1170–1175.

5. *Video as a Basic Service of LTE Networks: Mobile vMOS Defining Network Requirements.* Available: <http://www.huawei.com/minisite/4-5g/en/industryjsdc-j.html> (Accessed: 01.04.2017).

6. P.1201. Amendment 2, Parametric Non-intrusive Assessment of Audiovisual Media Streaming Quality Amendment 2: New Appendix III Use of ITU-T P.1201 for Non-adaptive, Progressive Download Type Media Streaming. *ITU-T*, 2013. 23 p.

7. **Essaili A.E., Schroeder D., Staehle D. et al.** Quality-of-Experience Driven Adaptive HTTP Media Delivery. *IEEE ICC 2013 – Communication QoS, Reliability and Modeling Symp.*, Budapest, 2013, Pp. 2480–2485.

Received 03.07.2017

12. **Hosek J., Ries M., Vajsar P. et al.** Mobile web QoE Study for Smartphones // IEEE Globecom Workshops. Atlanta, 2013. Pp. 1157–1161.

13. **Hosek J., Vajsar P., Nagy L. et al.** Predicting user QoE satisfaction in current mobile networks // IEEE Intern. Conf. on Communications. Sydney, 2014. Pp. 1088–1093.

14. **Cormen T.H., Leiserson, C.E., Rivest R.L., Stein C.** Introduction to Algorithms. 3rd. MIT Press and McGraw-Hill.

15. **Моисеев Н.Н., Иванилов Ю.П., Столярова Е.М.** Методы оптимизации. М.: Наука, 1978. 351 с.

16. Dash Industry Forum // URL: <https://github.com/Dash-Industry-Forum/dash.js/wiki> (Дата обращения: 01.04.2017).

8. **Pu W., Zou Z., Chen C.W.** Video Adaptation Proxy for Wireless Dynamic Adaptive Streaming over HTTP. *Proc. of 19th Intern. Packet Video Workshop*, Munich, 2012, Pp. 65–70.

9. **Chen Y.C., Chang J.W., Wei H.Y.** A Multi-level QoE Framework for Smartphone Video Streaming Applications. *2014 Globecom Workshops*, Austin TX, 2014, Pp. 225–230.

10. **Bui N., Valentin S., Widmer J.** Anticipatory Quality-Resource Allocation for Multi-user Mobile Video Streaming. *Proc. of the IEEE Conference on Computer Communications Workshops*, Hong Kong, 2015, Pp. 245–250.

11. **Khan A., Sun L., Jammeh E., Ifeakor E.** Quality of Experience-Driven Adaptation Scheme for Video Applications over Wireless Networks, *IET Communications*, 2010, No. 11, Pp. 1337–1347. DOI: 10.1049

12. **Hosek J., Ries M., Vajsar P. et al.** Mobile web QoE Study for Smartphones. *2013 IEEE Globecom Workshops*, Atlanta, 2013, Pp. 1157–1161.

13. **Hosek J., Vajsar P., Nagy L. et al.** Predicting user QoE Satisfaction in Current Mobile Networks. *Proc. of the IEEE Intern. Conference on Communications*, Sydney, 2014, Pp. 1088–1093.

14. **Cormen T.H., Leiserson, C.E., Rivest R.L., Stein C.** Introduction to Algorithms, 3rd., MIT Press and McGraw-Hill.

15. **Moiseev N.N., Ivanilov U.P., Stolyarova E.M.** *Metody optimizatsii [Optimization methods]*. Moscow: Nauka Publ., 1978, 351 p. (rus)

16. Dash Industry Forum. Available: <https://github.com/Dash-Industry-Forum/dash.js/wiki> (Accessed: 01.04.17).

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

ПАСТУШОК Игорь Анатольевич

PASTUSHOK Igor A.

E-mail: i.pastushok@vu.spb.ru

ТЮРЛИКОВ Андрей Михайлович

TURLIKOV Andrey M.

E-mail: turlikov@vu.spb.ru

DOI: 10.18721/JCSTCS.10303

УДК 004.4'6

ЭКСПЕРИМЕНТАЛЬНЫЕ АСПЕКТЫ ОЦЕНКИ ПРОПУСКНОЙ СПОСОБНОСТИ ПАМЯТИ КРУПНОМАСШТАБНЫХ СИСТЕМ С АРХИТЕКТУРОЙ CCNUMA

П.Д. Дробинцев, В.П. Котляров, А.В. Левченко

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

Рассмотрен подход к прогнозированию производительности широкого спектра научных приложений на современных перспективных архитектурах с глобально адресуемой памятью. Исследованы вопросы моделирования и прогнозирования пропускной способности памяти для приложений с гибридным параллелизмом MPI/OpenMP. Тест производительности HPCG использован для создания рабочей нагрузки, репрезентативной широкому спектру вычислительных и коммуникационных задач актуальных научных приложений. Проведены эксперименты по проверке модели на реальном кластере с общей памятью, имеющем архитектуру ccNUMA с 12 ТБ RAM и загруженном в режиме единого образа операционной системы, с целью определения границы размера задачи и демонстрации улучшенных показателей для целевой архитектуры по сравнению с базовой моделью. Модель позволит надежно оценить производительность современных и будущих систем ccNUMA, имеющих более 24 ТБ оперативной памяти на одном узле, и сравнить их экспериментальные результаты с другими проблемно-ориентированными архитектурами во всем мире.

Ключевые слова: оценка производительности; ccNUMA; HPCG; пропускная способность памяти; архитектура NUMA.

Ссылка при цитировании: Дробинцев П.Д., Котляров В.П., Левченко А.В. Экспериментальные аспекты оценки пропускной способности памяти крупномасштабных систем с архитектурой ccNUMA // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2017. Т. 10. № 3. С. 32–41. DOI: 10.18721/JCSTCS.10303

EXPERIMENTAL ASPECTS OF MEMORY BANDWIDTH FOR HPC SYSTEMS WITH CCNUMA ARCHITECTURE

P.D. Drobintsev, V.P. Kotlyarov, A.V. Levchenko

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

We have considered an approach to estimating the performance for a wide range of science applications calculated on modern HPC systems with globally addressed memory. Modeling and estimation of memory bandwidth have been examined for a set of applications with parallel structure based on MPI/OpenMP technology. The HPCG benchmark was used to create a workload representing a wide range of calculation and communication tasks in science applications. A set of experiments for checking the model on a real HPC system with globally addressed memory (ccNUMA architecture with 12 Tb of memory with single image of operating system installed) was conducted for estimating the size of the task and highlighting the benefits of

optimized model usage. The optimized model will allow to estimate the performance of modern and future systems developed based on the ccNUMA architecture which contains 24 Tb of memory in one node. The model will also allow to compare the results of NUMA systems with other modern HPC architectures.

Keywords: benchmarking; ccNUMA; HPCG; memory bandwidth; NUMA architecture.

Citation: Drobintsev P.D., Kotlyarov V.P., Levchenko A.V. Experimental aspects of memory bandwidth for HPC systems with ccNUMA architecture. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems. 2017, Vol. 10, No. 3, Pp. 32–41. DOI: 10.18721/JCSTCS.10303

Введение

В современных условиях промышленного подхода к решению сложных задач моделирования для организации широкого спектра процессов цифрового управления производством, наукой, сервисом требуются огромные вычислительные мощности суперкомпьютеров. Для их эффективного использования и прогнозирования производительности реализации приложений на современных перспективных архитектурах суперкомпьютеров с глобально адресуемой памятью важны исследования пропускной способности памяти для приложений с гибридным параллелизмом MPI/OpenMP.

Особенности архитектуры ccNUMA

Современные системы с архитектурой Cache-Coherent Uniform Memory Access (ccNUMA) могут предоставить большее количество оперативной памяти на один многомашинный узел (макроузел) с одним образом операционной системы, чем это доступно в стандартном вычислительном кластере. Однако асимметричная природа архитектуры ccNUMA порождает ряд негативных эффектов неоднородности, таких, например, как образование «горячих точек» памяти (memory hot-spotting), варьирующаяся сложная многоуровневая структура задержки удаленного доступа и несовпадение моделей доступа к данным и фактического распределения данных в памяти [1–3]. Эти и другие факторы оказывают разнонаправленное воздействие на пропускную способность памяти макроузла ccNUMA, что порождает серьезные проблемы для научных приложений с интенсивным нерегулярным доступом к памяти. При этом определение величины пропускной способности памяти, исходя из теоретического максимума

для конкретной вычислительной системы, является сложной задачей [4]. Таким образом, существующий гипотетический прогноз пропускной способности памяти систем ccNUMA нередко неубедителен.

Главная цель настоящего исследования – надежная оценка производительности современных и будущих систем ccNUMA. В данной статье представлены только предварительные экспериментальные результаты оценки, моделирования и прогнозирования пропускной способности памяти ccNUMA. Тест производительности HPCG (High Performance Conjugate Gradients) использован для создания рабочей нагрузки с низким отношением вычислений к доступу к данным, что характерно для основных коммуникационных и вычислительных моделей современных научных приложений [5].

В работе посредством расширения существующей модели производительности HPCG спрогнозирована эффективная пропускная способность памяти реальной системы с глобально адресуемой памятью, оснащенной 12 ТБ логически неделимой оперативной памяти и загружаемой в режиме единого образа операционной системы. Эксперимент дал возможность сопоставить полученные результаты с результатами других проблемно-ориентированных высокопроизводительных архитектур и спрогнозировать пропускную способность памяти будущих систем ccNUMA. Также в работе продемонстрированы важные технические аспекты, связанные с запуском HPCG на системах с большим объемом памяти одного многомашинного узла.

Предпосылки исследования и существующие работы

Ключевые вопросы существующих пу-

бликаций, связанных с NUMA и HPCG, позволяют учесть широкий спектр проблем, порожденных архитектурой ccNUMA (сюда относятся (1) публикации, касающиеся моделирования производительности гибридного параллелизма MPI/OpenMP, (2) исследования эффектов NUMA, оказывающих влияние на производительность, и (3) публикации, непосредственно связанные с HPCG, в т. ч. описание референсной модели производительности HPCG).

Для гибридного параллелизма в [6] предложена модель, которая используется и как модель оценки пропускной способности памяти, и как модель оптимального распределения ядер. В работе [7] дана ценная информация о моделировании межпроцессорной пропускной способности памяти для анализа производительности различных потоков и локаций размещения данных.

Гибридный подход для разработки высокоуровневых моделей производительности крупномасштабных вычислительных систем, сочетающий математическое моделирование и дискретное случайное моделирование, представлен в работе [8]. В исследовании [9] показаны преимущества гибридного программирования MPI/OpenMP для крупномасштабных кластеров NUMA. Исследования по моделированию производительности коммуникационных задач и вычислений в гибридных приложениях MPI/OpenMP выполнены в работе [9].

Работы по HPCG известны с 2013 г., с момента первой реализации теста. В работе Dongarra et al. [5] дано описание разрешенных и запрещенных оптимизаций HPCG. Несколько исследований [10–16] описывают ранний опыт оптимизации HPCG на Tianhe-2, «Ангара» и Sunway TaihuLight System. В дополнение к практическому опыту эксплуатации теста существует базовая аналитическая модель производительности [4] HPCG, которая является основополагающей для настоящей работы. Модель позволяет оценить время выполнения главных вычислительных и коммуникационных задач для реализации метода симметричного Гаусса-Зейделя (SymGS), вычисления произведения разреженной матрицы на вектор (SpMV), вычисления скалярного произве-

дения векторов (DDOT), вычисления суммы векторов (WAXPY). Вместе с моделью процедур межузловых коммуникаций полная модель позволит надежно прогнозировать производительность HPCG.

Как следует из сказанного выше, основная ценность исследования состоит в том, что HPCG может внести ясность в вопрос сопоставления производительности ccNUMA с результатами множества проблемно-ориентированных архитектур, отличных от ccNUMA. Эволюционные аспекты и экспериментальное применение упомянутых работ являются вкладом настоящего исследования.

Особенности расширенной модели

Основное достижение настоящей работы – это экспериментальное предсказание пропускной способности памяти системы с архитектурой ccNUMA с использованием эталонной модели [4]. Известными проблемами производительности параллельных приложений на архитектуре ccNUMA являются: (1) локальность доступа к данным; (2) объем обмена данными между потоками; (3) эффективная пропускная способность памяти [17].

Большое значение имеет эффективная пропускная способность памяти, которая является параметром в модели всех вычислительных процедур HPCG. Вклад настоящей работы – в использовании гибридной версии HPCG, а не только версии с чистым MPI, как в модели [4]. Следует учитывать, что HPCG хорошо сбалансирован на уровне MPI, отсюда производительность реализации MPI выше, чем производительность гибридной версии MPI/OpenMP. Кроме того, OpenMP не обеспечивает поддержку ccNUMA. Несмотря на это, идея исследования заключается в том, что гибридная версия является дополнительным серьезным вызовом для многомашинных узлов с архитектурой ccNUMA, провоцируя появление ряда эффектов, критичных для производительности, таких как «горячие точки» памяти (memory hot-spotting).

В табл. 1 показан оценочный диапазон вариантов модели, которые были рассмотрены или имеют такую перспективу. В этой

Таблица 1

Сравнение базовой и расширенной версии модели

Возможности модели	Базовая	Расширенная
SYMGS (время выполнения)	Рассмотрено	+ BW
SpMV (время выполнения)	Рассмотрено	+ BW
WAXPB (время выполнения)	Рассмотрено	+ BW
DDOT (время выполнения)	Рассмотрено	+ BW
Allreduce, Halo (время выполнения)	Рассмотрено	+ Задержка выч. сети
Гибридный параллелизм (MPI/OpenMP)	Не рассмотрено	Рассмотрено
Пропускная способность памяти	Не рассмотрено	Рассмотрено
Задержка вычислительной сети	Не рассмотрено	Рассмотрено
Факторы NUMA	Не рассмотрено	Рассмотрено
Техники оптимизации HPCG	Не рассмотрено	Не рассмотрено

статье предложены пока только экспериментальные аспекты эффективной оценки пропускной способности памяти (BW).

Из негибридной модели известно полное время выполнения [4]:

$$Iter_{time(sec)} = MG + SpMV(depth = 0) + 3(DDOT + WAXPB).$$

Гибридная версия HPCG (MPI/OpenMP) более чувствительна к пропускной способности памяти, чем версия MPI, и может иметь большую производительность [18], особенно для случая системы с глобально адресуемой памятью. Для OpenMP модель времени выполнения предложена ранее в работе Wu et al. [19]:

$$Perf = (Re f_{MPI} + OMP) \times \frac{Total_{exec_time(sec)}}{Comp_{exec_time(sec)} + Comm_{exec_time(sec)}},$$

где OMP представляет модель внутриузловой производительности OpenMP:

$$OMP = T_{c1} + (BW_n - 1) \frac{T_{c2} - T_{c1}}{BW_2 - 1}.$$

Здесь использована формула OMP для моделирования времени выполнения приложения OpenMP на n ядрах, основанная на модели производительности для 1 и 2 ядер (T_c) и пропускной способности памяти (BW_n) [19]. Пропускную способность памяти можно получить из базовой модели [4] для каждого вычислительного ядра

HPCG следующим образом:

$$\begin{aligned} BW_{SYMGS}(Bytes / sec) &= \frac{(nx \times ny \times nz) / 2^{3 \times d} \times (20 + 20 \times 27)(Bytes)}{SYMGS_{exec_time(sec)}}, \\ BW_{SpMV}(Bytes / sec) &= \frac{(nx \times ny \times nz) / 2^{3 \times d} \times (20 + 20 \times 27)(Bytes)}{SpMV_{exec_time(sec)}}, \\ BW_{WAXPB}(Bytes / sec) &= \frac{(nx \times ny \times nz) / 2^{3 \times d} \times 24(Bytes)}{WAXPB_{exec_time(sec)}}, \\ BW_{DDOT}(Bytes / sec) &= \frac{(nx \times ny \times nz) / 2^{3 \times d} \times 16(Bytes)}{DDOT_{exec_time(sec)}}, \end{aligned}$$

где $SYMGS$ является наиболее затратным [20].

Несмотря на то, что все вычислительные процедуры были смоделированы исчерпывающим образом, остаются важные параметры, полученные эмпирически. К их числу, помимо пропускной способности памяти, относится задержка вычислительной сети. Влияние задержки на прогноз производительности HPCG оценивается авторами базовой модели [4] как незначительное. В работе для эмпирической оценки латентности вычислительной сети использован модуль KNEM ядра Linux, обеспечивающий высокопроизводительную внутриузловую MPI-связь для больших сообщений [21].

Таблица 2

Варианты конфигурации

Особенности архитектуры	Базовый сервер	Макроузлы с единым образом ОС		
		Минимальный	Средний	Максимальный
RAM	188 ГБ	752 ГБ	3 ТБ	12 ТБ
Узлы NUMA	6	24	96	384
Плата/Сокет/Ядра	1/3/48	4/12/192	16/48/768	64/192/3072

Экспериментальные результаты

Поскольку системы ccNUMA с объемом памяти более 3 ТБ являются достаточно редкими, и кажется сложным получить набор различных гигантских многомашинных узлов, мы используем систему ccNUMA в следующих конфигурациях (табл. 2).

Для более глубокого изучения проблем, связанных с NUMA, проведены эксперименты с гибридной версией HPCG. Данная версия была запущена на макроузлах с 188 ГБ RAM (48 ядер) с объединением памяти макроузла до 3 ТБ RAM (768 ядер) и с последующей интеграцией в один макроузел с объемом памяти до ≈ 12 ТБ RAM (3072 ядра) на заключительном этапе. Базовый для многомашинного узла сервер основан на процессоре AMD Opteron 6380, межсоединение имеет топологию 3D Torus.

Используем ядро Linux 4.11 с набором патчей для поддержки драйвера Block Transfer Engine для контроллеров узлов NumaChip.

HPCG был скомпилирован с оптимизированной библиотекой *libgomp*, которая поддерживает локальное хранилище потоков (thread-local storage – TLS) для данных с числом потоков более 1024. Отдельный стек размером до 2 ГБ выделялся под каждый поток HPCG. Генерация инструкций для предварительной выборки памяти использовалась для повышения производительности циклов, которые обращаются к большим массивам. Нагрузка балансировалась для повышения эффективности приложения OpenMP, распределяя потоки через все доступные узлы NUMA, используя FPU и уменьшая нагрузку на интерфейс памяти и кэш третьего уровня. Генерация инструк-

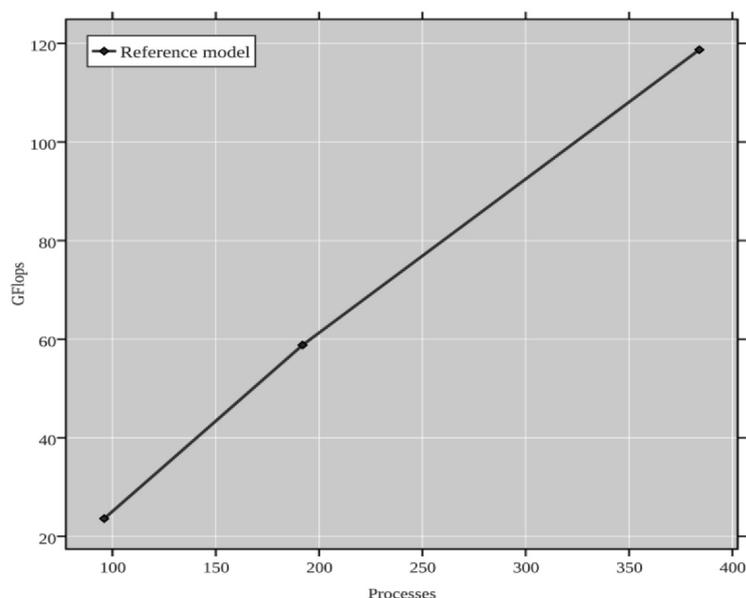


Рис. 1. Прогноз производительности HPCG, основанный на использовании базовой модели

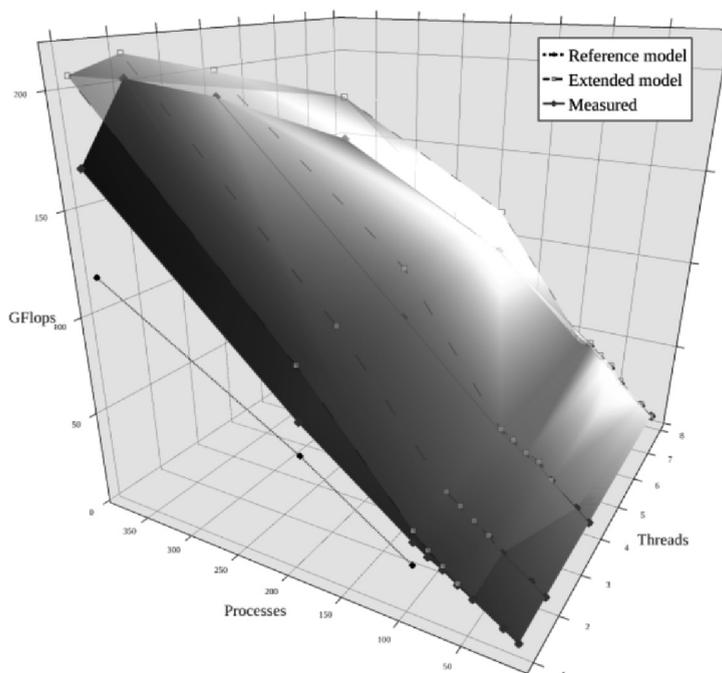


Рис. 2. Прогнозируемые и фактически полученные результаты HPCG на целевой крупномасштабной системе с 12 ТБ RAM

ций для предварительной выборки памяти использовалась для повышения производительности циклов обработки больших массивов. Результаты моделирования с помощью базовой модели приведены на рис. 1.

На рис. 2 сделанные прогнозы сопо-

ставлены с фактическими результатами гибридного HPCG на макроузле с 12 ТБ оперативной памяти. Кроме того, приведены результаты сравнения с эталонной моделью. В отличие от результатов работы [4], гибридный вариант HPCG нелинейно

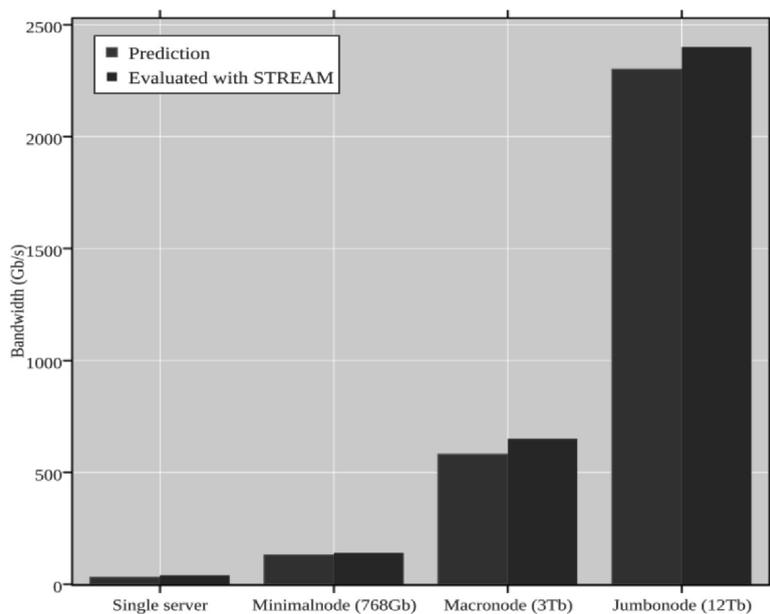


Рис. 3. Результаты теста STREAM и прогноз

масштабируется в системе ccNUMA; неравномерность системы приводит к разделению поверхности, причины которого исследуются. На рис. 3 приведено сравнение смоделированной пропускной способности памяти макроузла и результатов STREAM Benchmark.

В прогнозах производительности будущих систем ccNUMA с объемом памяти более 24 ТБ можно предположить, что пропускная способность памяти будет оставаться узким местом производительности HPCG. При крупномасштабном запуске HPCG с максимальным размером задачи около 7 ТБ памяти оказывается занято, причем прогнозируется пропорционально высокое потребление памяти, поскольку будущие системы ccNUMA будут иметь по крайней мере 4 ГБ RAM на ядро. Задержка вычислительной сети будет расти. Основываясь на данных моделирования, ожидается производительность не менее 400 GFlops для макроузла с 24 ТБ оперативной памяти.

В отношении существующих машин без ccNUMA, HPCG становится единой метрикой для сравнения различных проблемно-ориентированных архитектур и уменьшает разрыв в производительности, созданный оценками, сделанными на основе LINPACK. Например, экспериментальная система ccNUMA демонстрирует удовлетворительную производительность HPCG по сравнению с результатами технического отчета [15] для суперкомпьютера Sunway TaihuLight [22], предполагая, что память ccNUMA медленнее по сравнению с нынешним лидером TOP500.

Заключение

В статье представлен экспериментальный подход к оценке пропускной способности памяти современных систем ccNUMA. HPCG Benchmark использован для создания рабочей нагрузки, сопоставимой с современными научными приложениями. Существующая модель производительности HPCG расширена за счет использования гибридного MPI/OpenMP и дополнена факторами, влияющими на пропускную способность памяти. В результате предсказана пропускная способность памяти реальной системы ccNUMA с 12 ТБ оперативной памяти. Используемый подход может применяться для сравнения текущих и будущих машин ccNUMA.

Показано (рис. 3), что расхождение между фактически полученными результатами STREAM Benchmark и выведенными из эталонных моделей составляет до 12 %. Хотя вся программная среда была оптимизирована в широких масштабах (ядро Linux, gcc, libgomp, etc.), однако значительные оптимизации HPCG возможны. В ближайшем будущем планируется сосредоточиться на реализации существующих оптимизаций HPCG для случая ccNUMA, поскольку повышение производительности HPCG улучшит производительность реальных приложений [5].

В частности будет уточнена модель локализации кэш-памяти с помощью новой техники оптимизации HPCG, предложенной в статье [16]. Среди других улучшений предполагается заменить формат хранения матрицы CSR на упрощенный SELLPACK для ядер SpMV и SYMGS [10, 16]

Таблица 3

Оценка результатов

Техники оптимизации	Ожидаемые результаты
Раскраска вдоль двух областей XY одновременно в SYMGS [16]	Улучшение производительности SYMGS в три раза
Смена формата хранения (ELLPACK вместо CSR) [16]	Ускорение на 5 % в SYMGS и SPMV
Перераспределение данных [23]	Улучшение производительности для больших задач HPCG

(табл. 3 иллюстрирует ожидаемое ускорение). В работе [23] представлена новая модель перераспределения данных, позволяющая снизить издержки удаленного доступа к памяти для приложений с ин-

тенсивным вычислением с большими размерами проблемы. Наконец, в ближайшем будущем мы планируем предложить модель задержки вычислительной сети для многомашинного узла.

СПИСОК ЛИТЕРАТУРЫ

1. **Li T., Ren Y., Yu D., Jin S., Robertazzi T.** Characterization of input/output bandwidth performance models in NUMA architecture for data intensive applications // Proc. of the 42nd Internat. Conf. on Parallel Processing. 2013. Pp. 369–378.
2. **Diener M., Cruz E.H., Navaux P.O.** Modeling memory access behavior for data mapping // Internat. Journal of High Performance Computing Applications. 2016. URL: <http://hpc.sagepub.com/content/early/2016/04/13/1094342016640056.abstract>
3. **Zeng D., Zhu L., Liao X., Jin H.** A Data-Centric Tool to Improve the Performance of Multithreaded Program on NUMA. Springer International Publishing, 2015. Pp. 74–87 // URL: http://dx.doi.org/10.1007/978-3-319-27140-8_6
4. **Marjanović V., Gracia J., Glass C.W.** Performance modeling of the HPCG benchmark // High Performance Computing Systems. Performance Modeling, Benchmarking, and Simulation: 5th Internat. Workshop. New Orleans, LA, USA, 2014. Revised Selected Papers. Springer International Publishing, 2015. Pp. 172–192. URL: http://dx.doi.org/10.1007/978-3-319-17248-4_9
5. **Dongarra J., Heroux M.A., Luszczek P.** High-performance conjugate-gradient benchmark: A new metric for ranking high-performance computing systems // Internat. Journal of High Performance Computing Applications. 2016. Vol. 30. No. 1. Pp. 3–10. URL: <http://dx.doi.org/10.1177/1094342015593158>
6. **Wang W., Davidson J.W., Soffa M.L.** Predicting the memory bandwidth and optimal core allocations for multi-threaded applications on large-scale NUMA machines // Proc. of the IEEE Internat. Symp. on High Performance Computer Architecture. 2016. Pp. 419–431.
7. **Luo H., Brock J., Li P., Ding C., Ye C.** Compositional model of coherence and NUMA effects for optimizing thread and data placement // Proc. of the IEEE Internat. Symp. on Performance Analysis of Systems and Software. 2016. Pp. 151–152.
8. **Pllana S., Benkner S., Xhafa F., Barolli L.** Hybrid performance modeling and prediction of large-scale computing systems // Proc. of the Internat. Conf. on Complex, Intelligent and Software Intensive Systems, 2008. Pp. 132–138.
9. **Adhianto L., Chapman B.** Performance Modeling of Communication and Computation in Hybrid MPI and OpenMP Applications // Proc. of the 12th Internat. Conf. on Parallel and Distributed Systems. 2006. Vol. 2. Pp. 6.
10. **Zhang X., Yang C., Liu F., Liu Y., Lu Y.** Optimizing and scaling HPCG on Tianhe-2: Early experience // 14th Internat. Conf. on Algorithms and Architectures for Parallel Processing. Dalian, China, 2014. Proceedings, Part I. Springer International Publishing, 2014. Pp. 28–41. URL: http://dx.doi.org/10.1007/978-3-319-11197-1_3
11. **Chen C., Du Y., Jiang H., Zuo K., Yang C.** HPCG: Preliminary evaluation and optimization on Tianhe-2 CPU-only nodes // Computer Architecture and High Performance Computing. Proc. of the IEEE 26th Internat. Symp. on. 2014. Pp. 41–48.
12. **Liu Y., Zhang X., Yang C., Liu F., Lu Y.** Accelerating HPCG on Tianhe-2: A hybrid CPU-MIC algorithm // Proc. of the 20th IEEE Internat. Conf. on Parallel and Distributed Systems. 2014. Pp. 542–551.
13. **Liu F., Yang C., Liu Y., Zhang X., Lu Y.** Reducing communication overhead in the high performance conjugate gradient benchmark on Tianhe-2 // Proc. of the 13th Internat. Symp. on Distributed Computing and Applications to Business, Engineering and Science, 2014, Pp. 13–18.
14. **Liu Y., Yang C., Liu F., Zhang X., Lu Y., Du Y. et al.** 623 Tflop/s HPCG run on Tianhe-2: Leveraging millions of hybrid cores // Internat. Journal of High Performance Computing Applications. 2016. Vol. 30. No. 1. Pp. 39–54. URL: <http://hpc.sagepub.com/content/30/1/39.abstract>
15. **Dongarra J.** Report on the Sunway Taihu-light System. Oak Ridge National Laboratory, Department of Electrical Engineering and Computer Science, University of Tennessee. Tech. Rep. UT-EECS-16-742. Jun. 2016.
16. **Agarkov A., Semenov A., Simonov A.** Optimized implementation of HPCG benchmark on supercomputer with “Angara” interconnect // Proc. of the 1st Russian Conf. on Supercomputing – Supercomputing Days 2015. Moscow State University, 2015. No. 1. Pp. 294–302. URL: <http://ceur-ws.org/Vol-1482/294.pdf>
17. **Yang R., Antony J., Rendell A.P.** A simple performance model for multithreaded applications executing on non-uniform memory access computers // High Performance Computing and Commu-

nications. 2009.

18. **Nakajima K.** Flat MPI vs. Hybrid: Evaluation of parallel programming models for preconditioned iterative solvers on “T2K Open Supercomputer” // Proc. of the Internat. Conf. on Parallel Processing Workshops. 2009. Pp 73–80.

19. **Wu X., Taylor V.** Performance modeling of hybrid MPI/OpenMP scientific applications on large-scale multicore cluster systems // Proc. of the IEEE 14th Internat. Conf. on Computational Science and Engineering. 2011. Pp. 181–190.

20. **Park J., Smelyanskiy M., Vaidyanathan K., Heinecke A., Kalamkar D.D., Liu X., et al.** Efficient shared-memory implementation of high-performance conjugate gradient benchmark and its application to unstructured matrices // Proc. of the Internat. Conf. for High Performance Com-

puting, Networking, Storage and Analysis. 2014. Pp. 945–955.

21. **Goglin B., Moreaud S.** Knem: A generic and scalable kernel-assisted intra-node mpi communication framework // J. Parallel Distrib. Comput. 2013. Vol. 73. No. 2. Pp. 176–188. URL: <http://dx.doi.org/10.1016/j.jpdc.2012.09.016>

22. **Fu H., Liao J., Yang J., Wang L., Song Z., Huang X., Yang C., et al.** The Sunway TaihuLight supercomputer: system and applications // Science China Information Sciences. 2016. Vol. 59. No. 7. Pp. 1–16. URL: <http://dx.doi.org/10.1007/s11432-016-5588-7>

23. **Zhang M., Gu N., Ren K.** Optimization of computation-intensive applications in cc-NUMA architecture // Proc. of the Internat. Conf. on Networking and Network Applications. 2016. Pp. 244–249.

Статья поступила в редакцию 02.08.2017

REFERENCES

1. **Li T., Ren Y., Yu D., Jin S., Robertazzi T.** Characterization of input/output bandwidth performance models in NUMA architecture for data intensive applications. *Proc. of the 42nd International Conference on Parallel Processing*, 2013, Pp. 369–378.

2. **Diener M., Cruz E.H., Navaux P.O.** Modeling memory access behavior for data mapping. *International Journal of High Performance Computing Applications*. 2016. Available: <http://hpc.sagepub.com/content/early/2016/04/13/1094342016640056.abstract>

3. **Zeng D., Zhu L., Liao X., Jin H.** *A Data-Centric Tool to Improve the Performance of Multi-threaded Program on NUMA*. Springer International Publishing, 2015, Pp. 74–87. Available: http://dx.doi.org/10.1007/978-3-319-27140-8_6

4. **Marjanović V., Gracia J., Glass C.W.** Performance modeling of the HPCG benchmark. *High Performance Computing Systems. Performance Modeling, Benchmarking, and Simulation: 5th International Workshop, PMBS 2014*, New Orleans, LA, USA, 2014. Revised Selected Papers. Springer International Publishing, 2015, Pp. 172–192. Available: http://dx.doi.org/10.1007/978-3-319-17248-4_9

5. **Dongarra J., Heroux M.A., Luszczek P.** High-performance conjugate-gradient benchmark: A new metric for ranking high-performance computing systems. *International Journal of High Performance Computing Applications*, 2016, Vol. 30, No. 1, Pp. 3–10. Available: <http://dx.doi.org/10.1177/1094342015593158>

6. **Wang W., Davidson J.W., Soffa M.L.** Predicting the memory bandwidth and optimal core allocations for multi-threaded applications on large-scale

NUMA machines. *Proc. of the IEEE International Symposium on High Performance Computer Architecture*, 2016, Pp. 419–431.

7. **Luo H., Brock J., Li P., Ding C., Ye C.** Compositional model of coherence and NUMA effects for optimizing thread and data placement. *Proc. of the IEEE International Symposium on Performance Analysis of Systems and Software*, 2016, Pp. 151–152.

8. **Pilana S., Benkner S., Xhafa F., Barolli L.** Hybrid performance modeling and prediction of large-scale computing systems. *Proc. of the International Conference on Complex, Intelligent and Software Intensive Systems*, 2008, Pp. 132–138.

9. **Adhianto L., Chapman B.** Performance Modeling of Communication and Computation in Hybrid MPI and OpenMP Applications. *Proc. of the 12th International Conference on Parallel and Distributed Systems*, 2006, Vol. 2, Pp. 6.

10. **Zhang X., Yang C., Liu F., Liu Y., Lu Y.** Optimizing and scaling HPCG on Tianhe-2: Early experience. *14th International Conference on Algorithms and Architectures for Parallel Processing*. Dalian, China, 2014. Proceedings, Part I. Springer International Publishing, 2014, Pp. 28–41. Available: http://dx.doi.org/10.1007/978-3-319-11197-1_3

11. **Chen C., Du Y., Jiang H., Zuo K., Yang C.** HPCG: Preliminary evaluation and optimization on Tianhe-2 CPU-only nodes. *Proc. of the IEEE 26th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*, 2014, Pp. 41–48.

12. **Liu Y., Zhang X., Yang C., Liu F., Lu Y.** Accelerating HPCG on Tianhe-2: A hybrid CPU-MIC algorithm. *Proc. of the 20th IEEE International Con-*

ference on Parallel and Distributed Systems, 2014, Pp. 542–551.

13. **Liu F., Yang C., Liu Y., Zhang X., Lu Y.** Reducing communication overhead in the high performance conjugate gradient benchmark on Tianhe-2. *Proc. of the 13th International Symposium on Distributed Computing and Applications to Business, Engineering and Science*, 2014, Pp. 13–18.

14. **Liu Y., Yang C., Liu F., Zhang X., Lu Y., Du Y., et al.** 623 Tflop/s HPCG run on Tianhe-2: Leveraging millions of hybrid cores. *International Journal of High Performance Computing Applications*, 2016, Vol. 30, No. 1, Pp. 39–54. Available: <http://hpc.sagepub.com/content/30/1/39.abstract>

15. **Dongarra J.** *Report on the Sunway TaihuLight System*. Oak Ridge National Laboratory, Department of Electrical Engineering and Computer Science, University of Tennessee, Tech. Rep. UT-EECS-16-742, Jun 2016.

16. **Agarkov A., Semenov A., Simonov A.** Optimized implementation of HPCG benchmark on supercomputer with “Angara” interconnect. *Proceedings of the 1st Russian Conference on Supercomputing – Supercomputing Days 2015*, Moscow State University, 2015, No. 1, Pp. 294–302. Available: <http://ceur-ws.org/Vol-1482/294.pdf>

17. **Yang R., Antony J., Rendell A.P.** A simple performance model for multithreaded applications executing on non-uniform memory access computers. *High Performance Computing and Communications*, 2009.

18. **Nakajima K.** Flat MPI vs. Hybrid: Evaluation

of parallel programming models for preconditioned iterative solvers on “T2K Open Supercomputer”. *Proc. of the International Conference on Parallel Processing Workshops*, 2009, Pp 73–80.

19. **Wu X., Taylor V.** Performance modeling of hybrid MPI/OpenMP scientific applications on large-scale multicore cluster systems. *Proc. of the IEEE 14th International Conference on Computational Science and Engineering*, 2011, Pp. 181–190.

20. **Park J., Smelyanskiy M., Vaidyanathan K., Heinecke A., Kalamkar D.D., Liu X., et al.** Efficient shared-memory implementation of high-performance conjugate gradient benchmark and its application to unstructured matrices. *Proc. of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2014, Pp. 945–955.

21. **Goglin B., Moreaud S.** Knem: A generic and scalable kernel-assisted intra-node mpi communication framework. *J. Parallel Distrib. Comput.*, 2013, Vol. 73, No. 2, Pp. 176–188. Available: <http://dx.doi.org/10.1016/j.jpdc.2012.09.016>

22. **Fu H., Liao J., Yang J., Wang L., Song Z., Huang X., Yang C., et al.** The Sunway TaihuLight supercomputer: system and applications. *Science China Information Sciences*, 2016, Vol. 59, No. 7, Pp. 1–16. Available: <http://dx.doi.org/10.1007/s11432-016-5588-7>

23. **Zhang M., Gu N., Ren K.** Optimization of computation-intensive applications in cc-NUMA architecture. *Proc. of the International Conference on Networking and Network Applications*, 2016, Pp. 244–249.

Received 02.08.2017

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

ДРОБИНЦЕВ Павел Дмитриевич

DROBINTSEV Pavel D.

E-mail: drob@ics2.ecd.spbstu.ru

КОТЛЯРОВ Всеволод Павлович

KOTLYAROV Vsevolod P.

E-mail: vpk@spbstu.ru

ЛЕВЧЕНКО Алексей Викторович

LEVCHENKO Aleksei V.

E-mail: 2@exp.org



Federated Conference on Computer Science and Information Systems

9-12 September, 2018
Poznań, Poland

Call for Event:

The FedCSIS multi-conference consists of a significant number of antecedent Events. However, the Call for Events is open to all areas of computer science and information systems and new proposals for associated Events (conferences, symposia, workshops, special sessions, project dissemination events, etc.) are cordially invited until **November 14, 2017**.

The proposals should be prepared according to requirements specified in the Call for Events on the conference website <http://www.fedcsis.org> and emailed in a single PDF file to secretariat@fedcsis.org

The Events will be selected based on the scientific merit and/or the relevance to ICT practice, the completeness and clarity of the proposal in addressing the requested information, the strength of the proposed Program Committee, and the capacity of the FedCSIS conference program.



Circuits and Systems for Receiving, Transmitting and Signal Processing



Dear Readers!

One of the Journal goals is promotion of results obtained by young researchers and scientists. This is the reason that a special section has been created in the current issue of the Journal. I mean the papers from our abroad Colleagues prepared and published in English. Authors of the papers have been graduated this June from our International Master program “Microelectronics of Telecommunications Systems”. The papers can be considered as essential parts of their Master thesis.

In general the papers are devoted to the modern topics of microelectronics and electrical engineering, in particular, to the design of

digitally controlled high frequency oscillators (“Low Power Digitally Controlled Oscillator in 65nm Technology”, authored by Omar Al-Karkhi and Denis Akhmetov) and to the design of analog-to-digital converters with error corrections (“Linearization of the Analog-to-Digital Converter for FPGA-based Direct Digital Receiver”, authored by Maja Krneta and Ivan Piatak).

As a conclusion I’d like to express our gratitude to the scientific supervisors and simultaneously co-authors Dr. D. Akhmetov and Dr. I. Piatak for valuable efforts which improved the manuscript quality.

Editor-in-Chief Prof. Dr. Alexander Korotkov

DOI: 10.18721/JCSTCS.10304

UDC 621.3.049.774.2

M. Krneta, I.M. Piatak

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

LINEARIZATION OF THE ANALOG-TO-DIGITAL CONVERTER FOR AN FPGA-BASED DIRECT DIGITAL RECEIVER

The goal of this work is to provide linearization of the analog-to-digital converter for an FPGA-based Direct Digital Receiver. One of the techniques for implementing that is the LUT-based calibration and correction technique inside the FPGA, which was used for obtaining better characteristics of SNDR, SFDR and INL of the FPGA-based parametrized 14-bit model of pipelined ADC. The experimental part consists of the implementation of the FPGA-based parametrized 14-bit model pipelined ADC using MATLAB/Simulink. Its calibration and LUT-based correction were performed using both MATLAB/Simulink and FPGA. The DDS generator was implemented inside the FPGA. Code was written in Verilog HDL. The values of the dynamic characteristics of the ADC (SNDR and SFDR) were obtained before and after calibration, and compared. The influence of the DC gain and that of the capacitors of the MDAC was taken into account and observed. The influence of changes in the DC gain and capacitor mismatch of MDAC inside the stage was also taken into account for both SNDR and SFDR, as well as for the static characteristic, INL, which was also observed.

Keywords: pipelined ADC, digital error calibration and correction, MATLAB, SFDR, FPGA.

Citation: Krneta M., Piatak I.M. Linearization of the Analog-to-Digital Converter for an FPGA-based Direct Digital Receiver. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems. 2017, Vol. 10, No. 3, Pp. 44–52. DOI: 10.18721/JCSTCS.10304

РАЗРАБОТКА ЦИФРОВОЙ КОРРЕКЦИИ ХАРАКТЕРИСТИК АЦП ДЛЯ ПРИЕМНИКА ПРЯМОГО УСИЛЕНИЯ НА ОСНОВЕ ПЛИС

M. Krneta, I.M. Piatak

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

Описана линейаризация конвейерного аналого-цифрового преобразователя для цифровых приемников прямого преобразования. Используются цифровая калибровка и коррекция с применением таблицы соответствия, позволившие улучшить динамические характеристики АЦП. Разработана модель конвейерного АЦП разрядностью 14 бит в среде MATLAB с использованием результатов компьютерного моделирования. Рассмотрено влияние разброса номиналов конденсаторов и величин коэффициента усиления по постоянному току ОУ каскада АЦП на динамические характеристики. Для дальнейшего проведения эксперимента как система калибровки и коррекции на основе таблицы соответствия, так и модель АЦП реализованы с использованием ПЛИС. Результаты

эксперимента с высокой точностью совпадают с результатами компьютерного моделирования и свидетельствуют об эффективности предложенной калибровки и коррекции.

Ключевые слова: конвейерный АЦП; цифровая калибровка и коррекция; MATLAB; динамический диапазон; ПЛИС.

Ссылка при цитировании: Крнета М., Пятак И.М. Разработка цифровой коррекции характеристик АЦП для приемника прямого усиления на основе ПЛИС // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2017. Т. 10. № 3. С. 44–52. DOI: 10.18721/JCSTCS.10304

Introduction

The goal of this study is to provide the linearization of an analog-to-digital converter (ADC) for an FPGA-based direct digital receiver. FPGA stands for Field Programmable Gate Array. In [1] there is a clear correspondence between the sensitivity and signal-to-noise and distortion (SNDR) of the ADC, i. e., the more SNDR, the more sensitive the receiver is, because noise floor is lower. Moreover, spurious-free dynamic range (SFDR) is important, because the less harmonics at the ADC output, the less the probability of blocking the desired low signals by these harmonics and less unwanted received components. Therefore, the correction of the ADC characteristics and especially linearization are important for obtaining high characteristics of the receiver (in particular, sensitivity).

In order to perform the linearization of the ADC, at first it is necessary to provide an ADC model which will take into account some well-known nonlinearities, like non-ideal DC gain and mismatch of the multiplying digital-to-analog converter (MDAC) capacitors inside the pipelined ADC stage. In this work the ADC model inside the FPGA is realized by using a direct digital synthesizer (DDS generator). Instead of sine ROM in DDS, we are using ROM with samples from the MATLAB model with nonidealities. So we can introduce nonlinearities in hardware by changing the DC gain and capacitor mismatch in software in the model of the ADC (MATLAB/Simulink), without buying expensive instruments. In this work, the correction of the dynamic characteristics, SNDR and SFDR, of the ADC is performed using histogram calibration [2].

Literature overview

In the literature we can find results obtained by other calibration and correction techniques for the 14-bit pipelined ADC. For example, for the LMS-based technique at the sampling rate of 200 Msps, SNDR after correction equals 69.5 dB, SFDR 88.9 dB [3]. For the redundancy-based technique [4], the sampling frequency 100 MHz, SNDR is 85.89 dB, SFDR 102.8 dB. For the dither-based technique [5], sampling frequency 100 MHz, SNDR equals 76.56 dB. For the PN-sequence-based technique [6], 125 Msps, 3.5-bit stage architecture, SNDR is 80.86 dB, SFDR 102.27 dB. The combined foreground + background technique, described in [7], at 100 Msps and capacitor mismatch of 0.1 % gave SNDR of 76.8 dB, SFDR of 75.7 dB and INL ± 1.2 LSB. For the background technique, described in [8], at 100 Msps and input frequency of 21 MHz, SNDR is 73 dB, SFDR 91 dB, INL 1.3 LSB. The charge-sharing background technique [9] at 100 Msps, input frequency 30 MHz, gives SNDR of 73.1 dB, SFDR of 91 dB and INL 1.1/–1.0 LSB.

In this study the look-up table based (LUT-based) correction is used because it is independent from the architecture of the ADC, it is suitable for many different ADCs, it is easy to perform as IP-block for FPGA, it has low power and improves the linearity and accuracy. The disadvantage of this technique is the following: it requires industrial hardware.

Linearization of the ADC

The first task was to implement the parameterized FPGA-based DDS generator. The DDS code was written in Verilog hardware description language (HDL). Memory was initialized inside the FPGA with the external

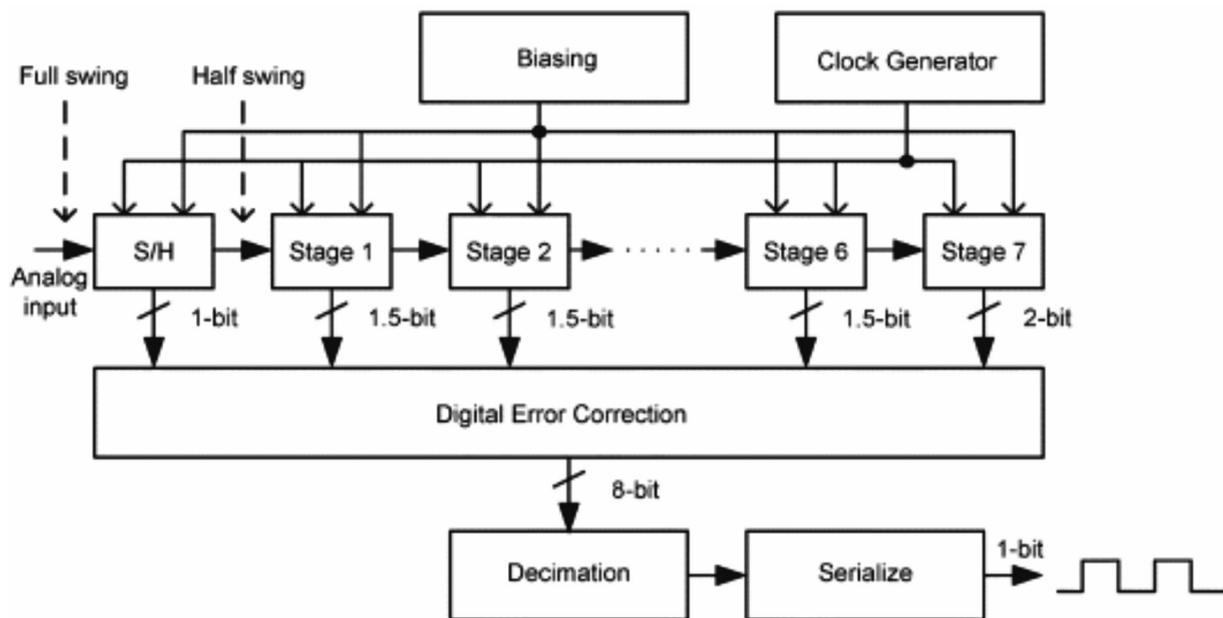


Fig. 1. Structure of the pipelined ADC [11]

file with sine coefficients. The possibility to initialize memory inside the FPGA with external coefficients is a key for building a model ADC and LUT-based correction. The example of the look-up table method is explained in detail in [10].

The Verilog project consists of one module. In this module memory is initialized with an external file, obtained from the non-ideal ADC MATLAB/Simulink model. In this project, hexadecimal sine coefficient values were used. Master clock frequency (MCLK) of the FPGA board has a value of 100 MHz. Sine output was 14 bits long.

Comparing the values from the MATLAB/Simulink ADC model and the values from the output of the DDS, it can be noticed that they are the same, which means that the memory can be initialized inside the FPGA with the external file, containing sine coefficients, and that the DDS generator works, which is exactly what had to be proven.

The next task was the implementation of the parameterized model of the FPGA-based pipelined ADC using MATLAB/Simulink. The structure of the pipelined ADC is shown in Fig. 1.

Fig. 2 shows the MATLAB/Simulink model

of the 14-bit pipelined ADC. Its calibration and correction in this part of the study were also done in MATLAB.

The dynamic characteristics SNDR and SFDR of the ADC have been measured. The DC gain of the ideal 14-bit pipelined ADC is 84 dB, the values of the MDAC capacitors C_s and C_f are equal to 1. The sampling frequency equals 100 MHz. Input frequency is 3 MHz.

First of all, the characteristics of the non-calibrated ADC have been obtained. After that, the calibration was done. The SNDR and SFDR values were recorded for the values of the DC gain of 54 and 84 dB. The value of 54 dB is used because it is enough for obtaining good dynamic characteristics using low power op-amps; after the LUT-based correction they are close to ideal values, which will be proven in the next steps. The same procedure was repeated for the fixed DC gain of the ideal 14-bit ADC (84 dB), but in this case the value of the capacitor C_f was changed. The value dC_f was added to the value of the C_f of the ideal 14-bit pipelined ADC (equals 1). The step size of the dC_f is 0.0002, the range of the dC_f values is from 0.001 to 0.002 (change of C_f from 0.1 % to 0.2 %). Those dC_f values are used in order to see the changes of SNDR and SFDR close

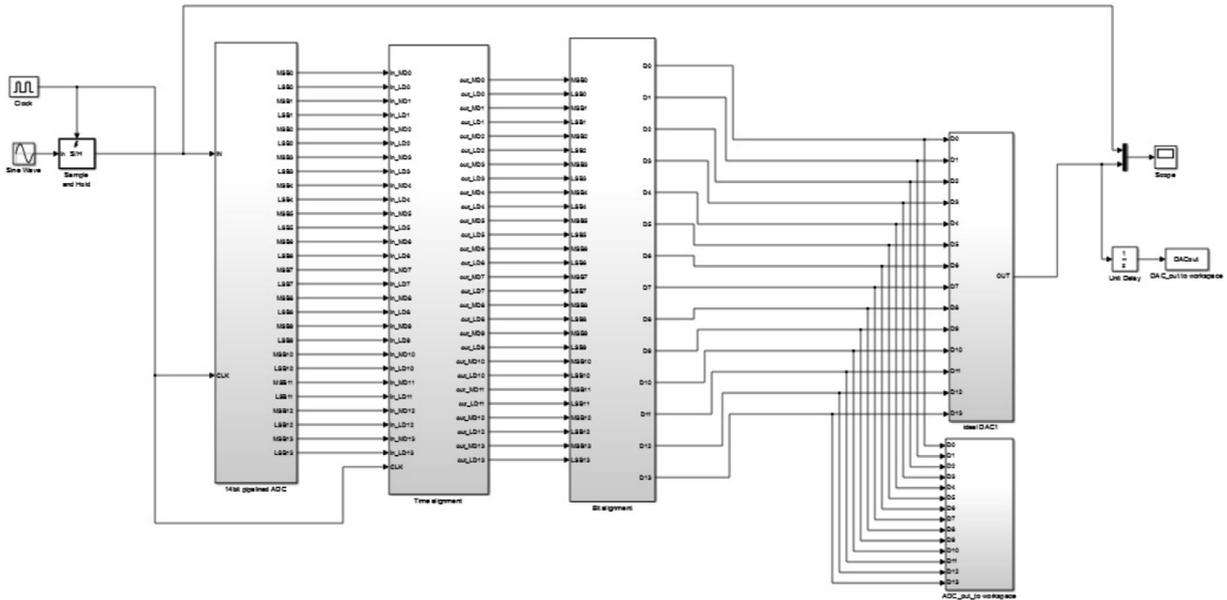


Fig. 2. MATLAB/Simulink model of the 14-bit pipelined ADC

to the down limit of 0.1 %. With this value it is impossible to make a mismatch because of technological operations during manufacturing. After that, the samples from MATLAB are used for initializing the DDS generator memory in the FPGA and after that the sine values obtained at the output of the DDS (Xilinx ISE) are loaded back to MATLAB. They are used in the calibration engine. Table lists the results obtained for the values of DC gain of 84 dB and 54 dB and for the capacitor mismatch of

0 and 0.14 %, before and after correction, using both MATLAB and FPGA. The value of 0.14 % is used as an example between 0.1 and 0.2 %. It can be noticed that the results from MATLAB and FPGA are pretty similar.

The graphs represented in the Fig. 3 and Fig. 4 show the comparison of both characteristics, with MATLAB- and FPGA-obtained values. From their similarity, we can conclude that it is possible to implement the parametrized model of FPGA-based pipelined

MATLAB and FPGA results of SNDR and SFDR for DC gain of 54 and 84 dB and dC_f of 0 and 0.14 %

		$dC_f = 0 \%$		$dC_f = 0.14 \%$	
		84 dB	54 dB	84 dB	54 dB
MATLAB before	SNDR	83.62	57.44	70.84	56.04
	SFDR	87.97	58.12	71.64	56.72
FPGA before	SNDR	83.62	57.44	N/A	N/A
	SFDR	87.97	58.12	N/A	N/A
MATLAB after	SNDR	82.13	77.24	82.03	76.36
	SFDR	90.35	98.76	95.71	97.95
FPGA after	SNDR	82.12	77.21	N/A	N/A
	SFDR	90.28	93.45	N/A	N/A

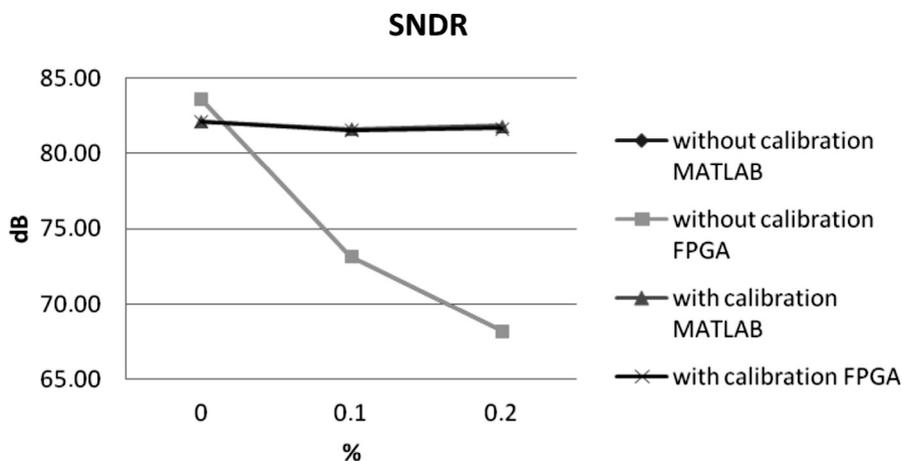


Fig. 3. SNDR characteristic for MATLAB/FPGA values, before and after calibration, DC gain 84 dB

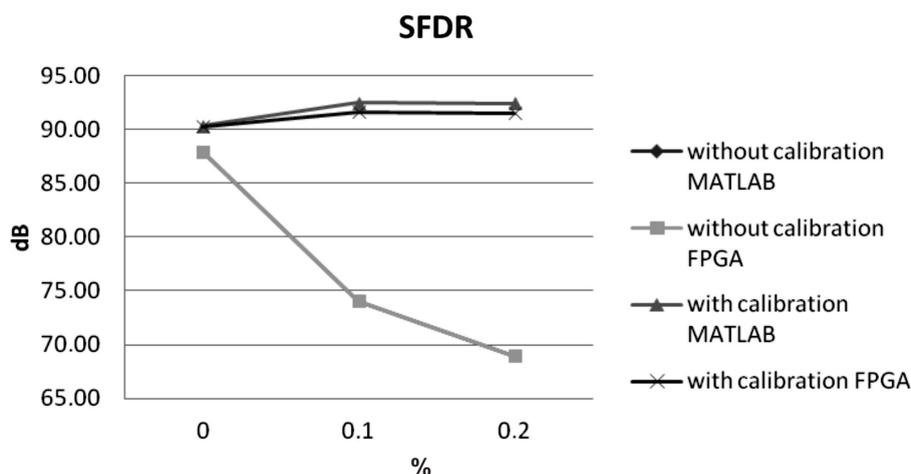


Fig. 4. SFDR characteristic for MATLAB/FPGA values, before and after calibration, DC gain 84 dB

ADC using MATLAB/Simulink.

The results show that the dynamic characteristics of the ADC, SNDR and SFDR depend on the changes of the DC gain, as well as on the changes of the MDAC capacitors. After the calibration, the characteristics became better, i. e., the values of the SNDR and SFDR increased, in comparison with their characteristics before the calibration. In view of this, it can be concluded that the LUT-based calibration improves the dynamic characteristics of the ADC and thus its linearization.

The next step is the implementation of LUT-correction of the parametrized pipelined model ADC, which is done inside the FPGA,

using MATLAB. The code was written in Verilog HDL.

The module contains DDS generator and LUT. For both, the memory is initialized with external file, containing sine samples. Samples for LUT are obtained from the LUT of correction in MATLAB/Simulink. Once again, sampling frequency equals 100 MHz. Input frequency is 3 MHz. Again, the dynamic characteristics, SNDR and SFDR, have been observed, as well as the behaviour of the static characteristic INL. The calibration and correction are done for the DC gain of 54 dB and for the capacitor mismatch of 0, 0.1, 0.12, 0.14, 0.16, 0.18 and 0.2 %.

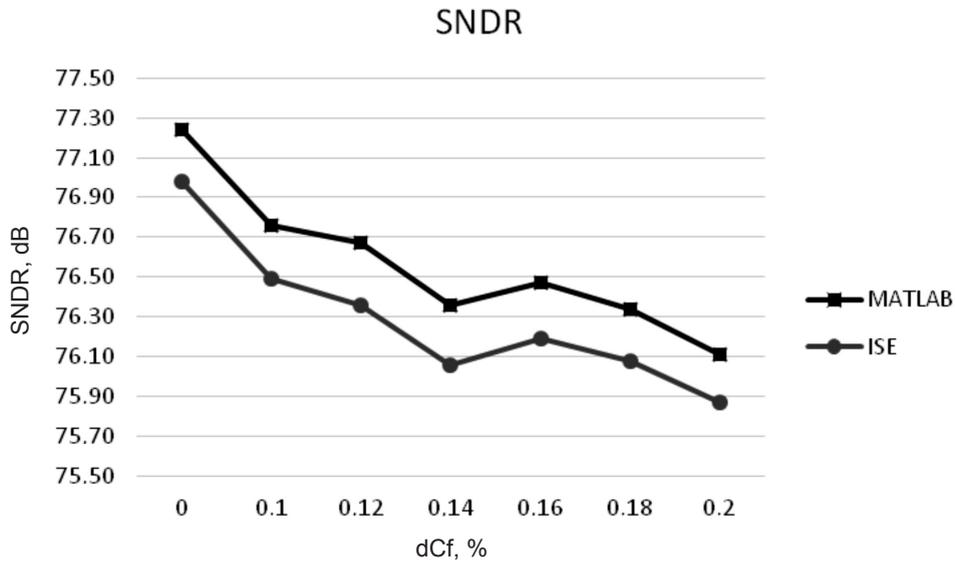


Fig. 5. SNDR after corrections in MATLAB and FPGA

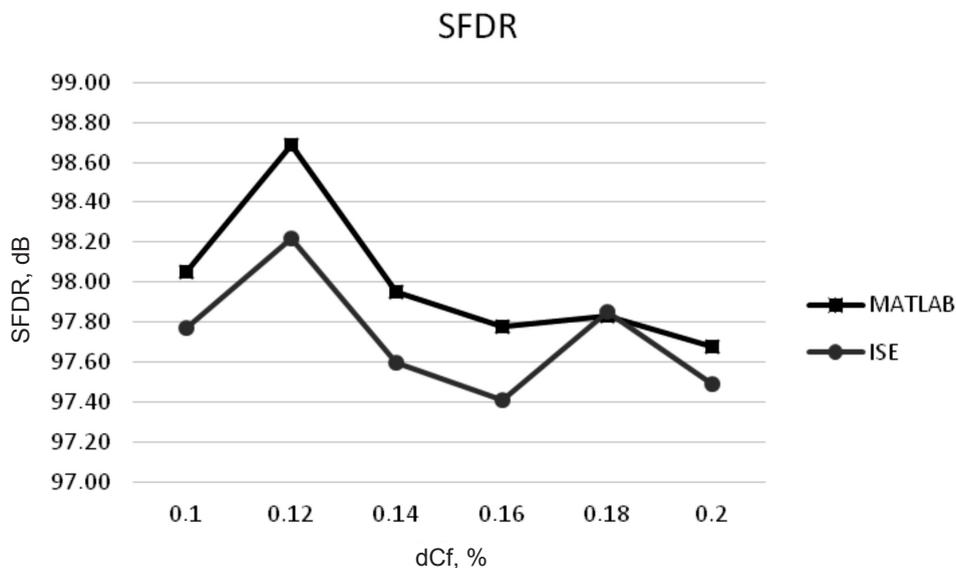


Fig. 6. SFDR after corrections in MATLAB and FPGA

The obtained results are compared with the results obtained from the LUT-based correction in MATLAB. Fig. 5 shows the values of SNDR and Fig. 6 shows the values of SFDR.

The spectra of the signal with capacitor mismatch after corrections in MATLAB and FPGA are shown in Fig. 7.

It can be seen that they are similar, which leads to the conclusion that LUT-based correction done inside the FPGA using

MATLAB gives the result similar to the one obtained by MATLAB correction directly. If we consider the values, as an example let us take the capacitor mismatch of 0.14 % for DC gain of 54 dB. SNDR after MATLAB correction is 76.36 dB, after FPGA correction 76.06 dB. SFDR after MATLAB correction is 97.95 dB, after FPGA correction 97.60 dB. Therefore, it is proven that the LUT-based calibration and correction can be implemented

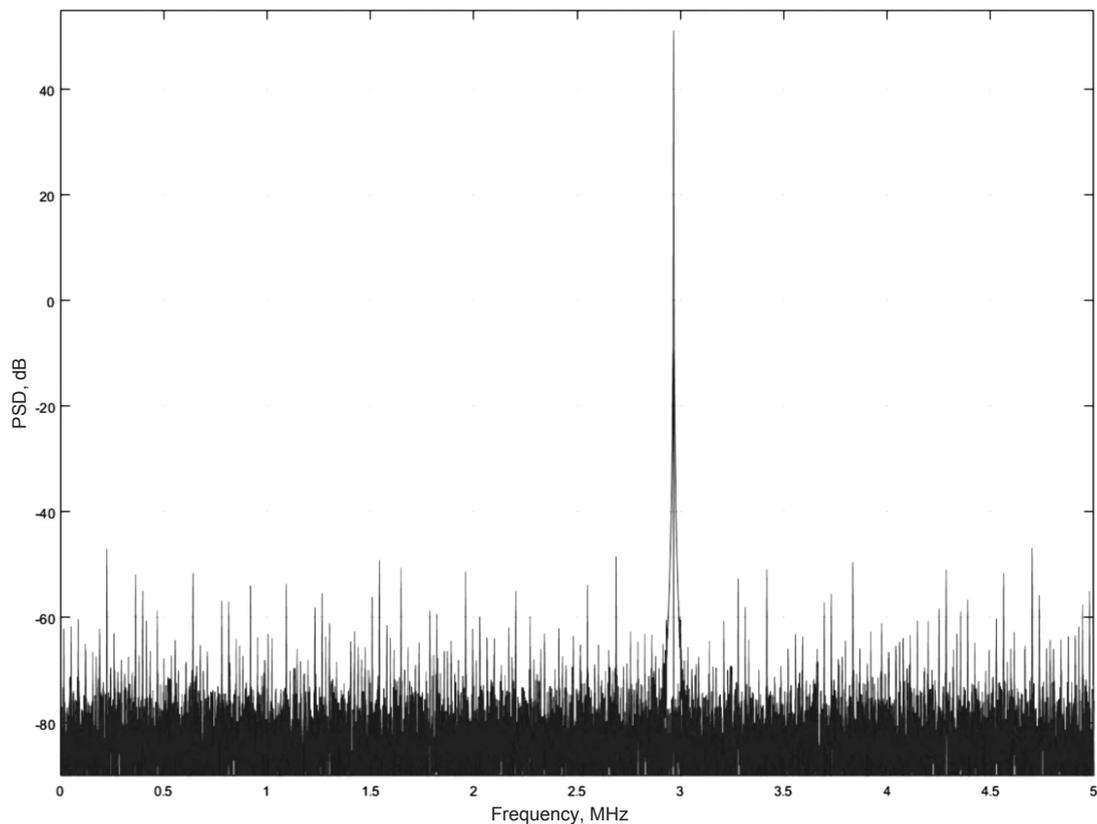


Fig. 7. Spectra after correction in MATLAB and FPGA ($dC_f = 0.14\%$)

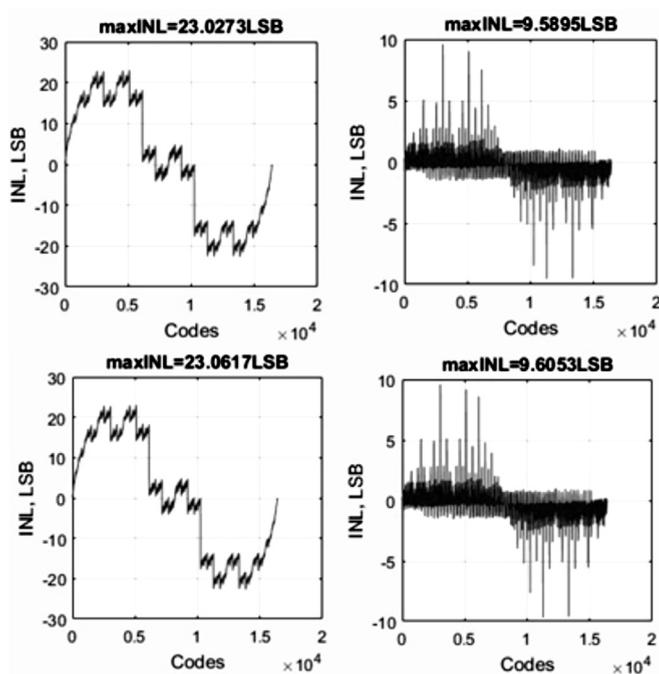


Fig. 8. INL before and after correction in MATLAB (up) and FPGA (down), 54 dB, 0.1 %

inside the FPGA with the same efficiency as in MATLAB.

The values of INL are obtained before and after correction inside the FPGA. The values were measured for 54 dB of DC gain and 0, 0.1 and 0.2 % of capacitor mismatch. In Fig. 8 the INL characteristic before and after correction using MATLAB (up) and FPGA (down) is represented.

There is a clear improvement of INL characteristic after LUT-based corrections, performed in both MATLAB and FPGA. INL values are around 2.5 times lower than before LUT-based correction.

Conclusions

In this study, it is shown that the nonlinearity of the ADC characteristics can be removed by LUT-based calibration and correction technique. It is also shown that SNDR, SFDR and INL depend on the changes of the DC gain, as well as on

the changes of the MDAC capacitors, i. e., capacitor mismatch inside the stage. Not only the simulations have been performed in this study, but the experiment as well, and the results obtained in the experiment are in good agreement with the results from simulation.

Using LUT-based calibration and correction inside the FPGA, the best results obtained in this study are the improvement of SNDR from 57.44 dB before correction to 76.98 dB after correction. For SFDR the biggest improvement obtained was from 58.12 dB to 98.22 dB. Therefore, with the LUT-based technique, SNDR can be improved for 20 dB and SFDR for 40 dB. After LUT-based correction INL is lower for around 2.5 times than before correction. Comparing the results obtained in this work with the results from the known techniques, it can be noticed that the LUT-based technique using FPGA can guarantee a similar level of improvement in SFDR, SNDR and INL, or even better.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. **Korotkov A.S.** Devices for receiving and processing signals. Microelectronic high-frequency devices of radio receivers of communication systems [Ustroystva priyema i obrabotki signalov. Mikroelektronnyye vysokochastotnyye ustroystva radiopriyemnikov sistem svyazi]. Available: <http://elib.spbstu.ru/dl/local/2762.pdf>. St. Petersburg, 2010. (rus)
2. **Okawara H.** Mixed Signal Lecture Series: DSP-Based Testing Fundamentals 18 - Histogram Method in ADC Linearity Test. *Verigy University Rev.*, Oct. 2009.
3. **Ye F., Zhang P., Yu B., Chen Ch., Zhu Yu, Ren J.** A 14-bit 200-MS/s time-interleaved ADC calibrated with LMS-FIR and interpolation filter. *International Conference of Electron Devices and Solid-State Circuits*, 2011, Pp. 1–4.
4. **Narula Sw., Pandey S.** High performance 14-bit pipelined redundant signed digit ADC. *Journal of Semiconductors*, Vol. 37, No. 3, Pp. 1–10.
5. **Cheng J.** Variable-amplitude dither-based digital background calibration algorithm for linear and high-order nonlinear error in pipelined ADCs. *Microelectronics Journal*, 2010, No. 41(7), Pp. 403–410.
6. **Fan K., Liu X., Lee A.** Digital calibration technique for A 14-bit 125-MS/s pipelined ADC using PN dithering. *International Symposium on Intelligent Signal Processing and Communications Systems*, 2011, Pp. 1–4.
7. **Wang X., Chen J., He L.** Modeling of A 14-bit, 100-MS/s Pipelined ADC with Digital Nonlinearity Calibration. *Proceedings of the 12th International Symposium on Integrated Circuits*, 2009, Pp. 208–211.
8. **Van de Vel H., et al.** A 1.2V 250mW 14-b 100 MS/s Digitally Calibrated Pipeline ADC in 90-nm CMOS. *Journal of Solid-State Circuits*, 2009, Pp. 74–75.
9. **Cai H., Li P.** A CMOS switch-capacitor 14-bit 100 Msps pipeline ADC with over 90 dB SFDR. *International Journal of Electronics*, 2013, Vol. 100, No. 1, Pp. 62–71.
10. **Gatti U., Gazzoli G., Maloberti F., Mazzoleni S.** A calibration technique for high-speed high-resolution A/D converters. *Proc. of the 3rd International Conference on Advanced A/D and D/A Conversion Techniques and Their Applications*, 1999, Pp. 168–171.
11. **Jiang Sh., Anh Do M., Yeo K.S., Lim W.M.** An 8-bit 200-MSample/s Pipelined ADC With Mixed-Mode Front-End S/H Circuit. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2008, Vol. 55, No. 6, Pp. 1430–1440.

Received / Статья поступила в редакцию 08.07.2017

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

KRNETA Maja

KPHEТА Мая

E-mail: mm.krmeta@gmail.com

PIATAK Ivan M.

ПЯТАК Иван Михайлович

E-mail: i.m.piatak@gmail.com

DOI: 10.18721/JCSTCS.10305

UDC 621.396

A LOW-POWER DIGITALLY CONTROLLED OSCILLATOR BASED ON 65-nm CMOS TECHNOLOGY

*D.B. Akhmetov*¹, *Omar J. Al-Karkhi*²

¹Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation;

² Ministry of Science and Technology, Baghdad-Iraq

This paper contains the research results related to the Master thesis about the design of Dual band Differential Digital Ring oscillator (DRO) in two stages, describing the lower power consumption, smaller area, lower phase noise, linear frequency range and better frequency stability with variation of applied voltage, in addition to the investigation of the temperature variation effect. We have proposed a circuit using the 65-nm CMOS process with Radio Frequency (RF) transistors and the output frequency digitally controlled (by 4-bit (coarse), 3-bit (fine) tuning) as control code, for the low-band frequency range [1.487–3.021 GHz] and power consumption of 0.359 mW @ 2.42 GHz, and for the high-band frequency range [3.5–6.98 GHz] and the power consumption of 1.86 mW @ 6.023 GHz, the band gap between two bands equal to 500 MHz, the phase noise about -84.4 dBc/Hz @ 1MHz, and the jitter value of 4.335 ps, with FOM equal to -156.5 dBc/Hz.

Keywords: 65-nm technology; DCO; positive feedback, power consumption; buffer; cadence software.

Citation: Akhmetov D.B., Al-Karkhi Omar J. A low-power digitally controlled oscillator based on 65-nm cmos technology. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems. 2017, Vol. 10, No. 3, Pp. 53–58. DOI: 10.18721/JCSTCS.10305

ГЕНЕРАТОР, УПРАВЛЯЕМЫЙ ЦИФРОВЫМ КОДОМ, С ПОНИЖЕННЫМ ЭНЕРГОПОТРЕБЛЕНИЕМ НА ОСНОВЕ 65 нм КМОП-ТЕХНОЛОГИИ

*Д.Б. Ахметов*¹, *Omar J. Al-Karkhi*²

¹Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация;

² Ministry of Science and Technology, Baghdad-Iraq

Изложены результаты исследований, связанных с магистерской диссертацией о конструкции двухканального дифференциального генератора цифровых колец (DRO) в два этапа, что отвечает за большую часть потребления энергии, меньшую площадь, более низкий фазовый шум, линейный частотный диапазон и лучшую стабильность частоты с изменением приложенного напряжения в дополнение к исследованию эффекта изменения температуры. Предложена схема двухдиапазонного генератора, разработанного с использованием параметров КМОП-технологии с разрешением 65 нм. Выходная частота генератора управляется в диапазонах 1,49–3,02 ГГц и 3,50–6,90 ГГц 4-разрядным цифровым кодом для грубой настройки и 3-разрядным – для точной настройки. Потребляемая мощность составляет 0,36 мВт и 1,86 мВт соответственно в каж-

дом диапазоне; уровень фазовых шумов – не более минус 84,4 дБс/Гц при отстройке 1 МГц; значение джиттера – не более 4,3 пс; комплексный параметр качества (FOM) – минус 156,5 дБс/Гц.

Ключевые слова: 65-нм технология; ИДК; положительная обратная связь; энергопотребление; буфер; программное обеспечение cadence.

Ссылка при цитировании: Ахметов Д.Б., Al-Karkhi Omar J. Генератор, управляемый цифровым кодом, с пониженным энергопотреблением на основе 65 нм КМОП-технологии // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. Т. 10. № 3. 2017. С. 53–58. DOI: 10.18721/JCSTCS.10305

The wireless systems are closely linked to the phenomenal success of the CMOS technology scaling that makes it possible to develop increasingly complicated systems on a single silicon chip while preserving performance and functionality at an ever lower cost, lower power consumption, smaller product size and an increased unity gain of CMOS transistors [1, 2]. Concerning wireless systems, LC circuits in controlled voltage oscillators and large-sized capacitors in filters used in frequency synthesizers are less suitable for scalability. In contrast, all-digital phase-locked loops (ADPLL) based on digitally controlled oscillators (DCO) can be easily integrated into the digital system and have much less dependence on temperature, process and voltage variations. The ADPLL architecture still entails significant levels of power consumption

and silicon area in a nanoscale CMOS, which allows for further power and cost reductions. The focus is on analyzing and tracking the advances in the DCO base depending on its performance level. As we knew that the digitally controlled oscillator has many types and topologies and because the Differential Ring Digital- Controlled-Oscillator (DRO) is better in frequency stability with power supply voltage variations and has lower phase noise, a DRO using the 65-nm technology in a two delay stages is proposed.

In this paper, the differential Digital-Controlled-Oscillator (DCO) with a reconfigurable delay cell in each two stages of a digitally controlled ring oscillator is proposed. The DCO operates in two bands: 1.40–3.02 GHz and 3.50–6.98 GHz in comparison with the architecture proposed in [3]. Each band has 128

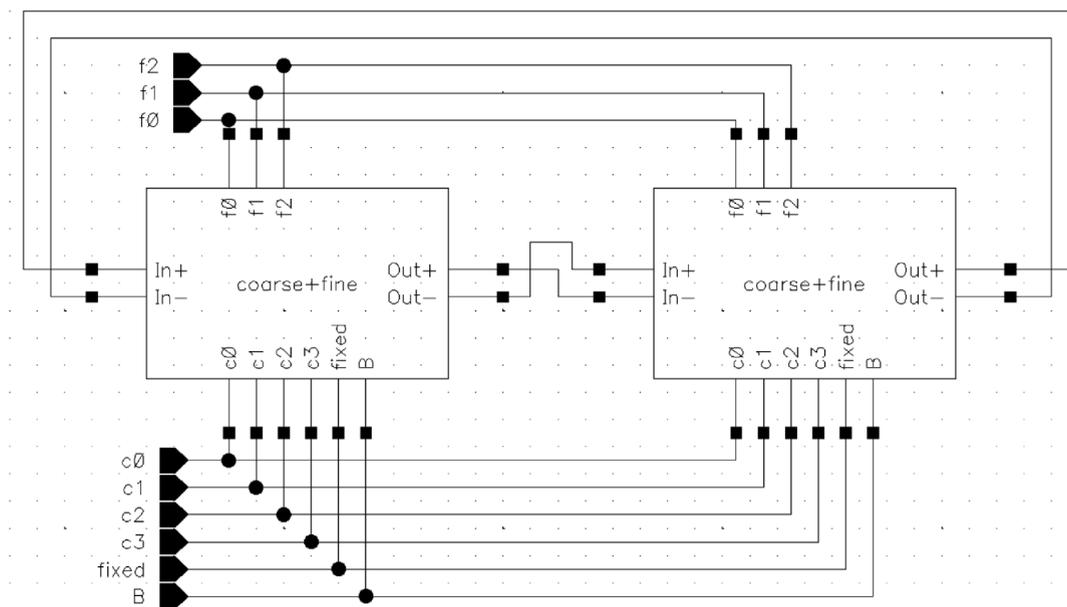


Fig. 1. DCO structure

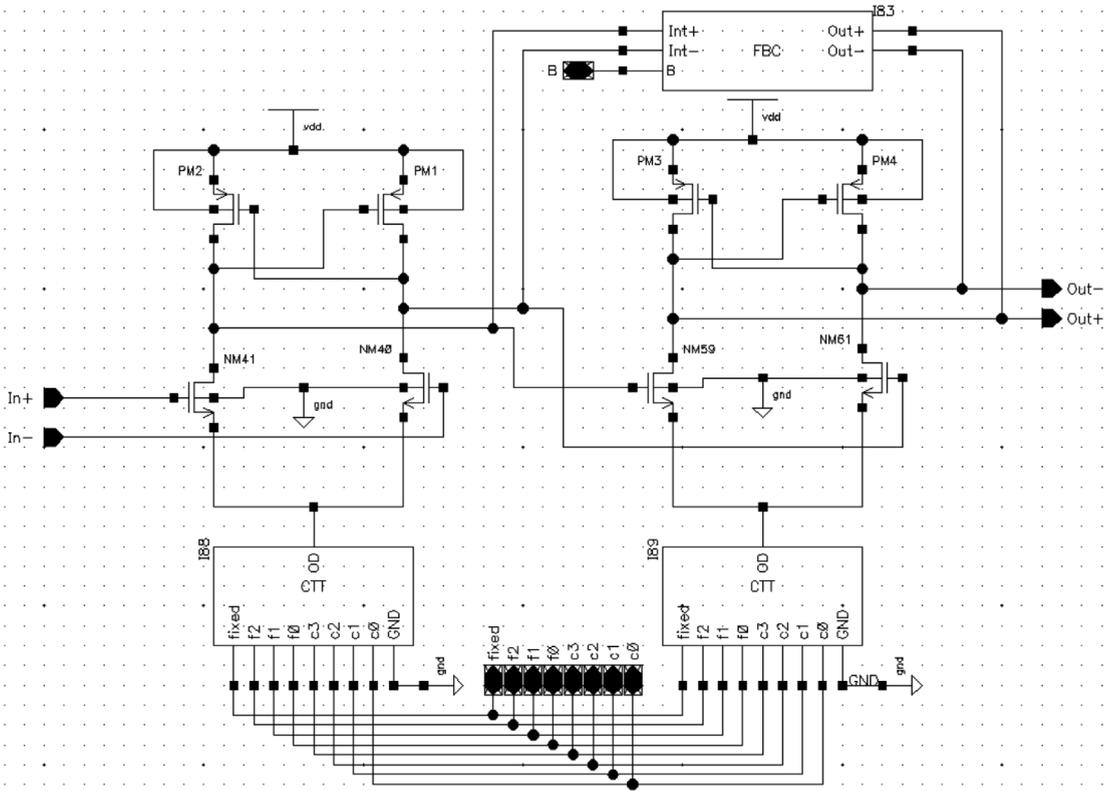


Fig. 2. Schematic of the delay element

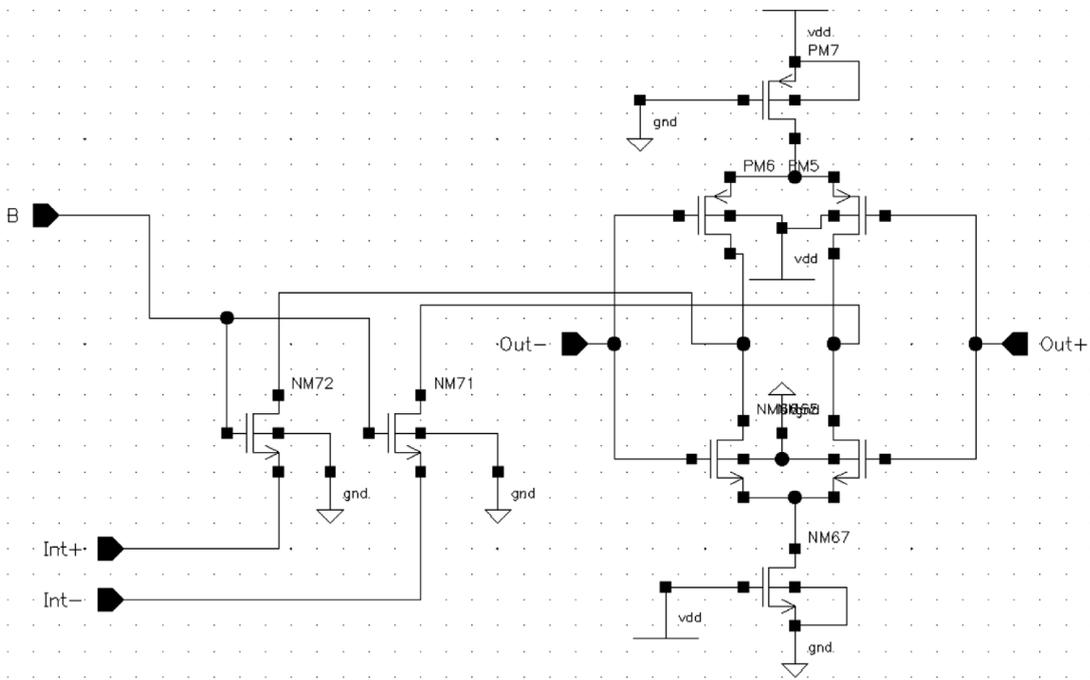


Fig. 3. Schematic of the «FBC»

sub-bands controlled by the 7-bit digital word, a new SCL-based delay cell is used which has a reconfigurable structure. Fig. 1 shows the proposed dual band DCO with a block diagram of delay elements.

Each delay element has the following pins: pin «B» defines the current band, pins «C0» – «C3» «F0», «F1», «F2» define the upper and lower bits of the digital control word. The schematic of the delay element is shown in Fig. 2

Contrary to the traditional trigger structure, a new feedback circuit «FBC» is included into the scheme. When the signal on pin «B»

is low, total propagation delay increases and the circuit operates at a lower frequency band. When the signal on pin «B» is kept high, the propagation delay decreases and frequency of oscillation rises. Block «CTT» contains a bank of transistors which are used to set the tail current of the differential pair. Schematic of the «FBC» is shown in Fig. 3.

Introducing a positive feedback into the trigger scheme allows using the same control signals for frequency control within each band. The frequency variations depending on the control word for low and high band are shown in Fig. 4 and Fig. 5, respectively.

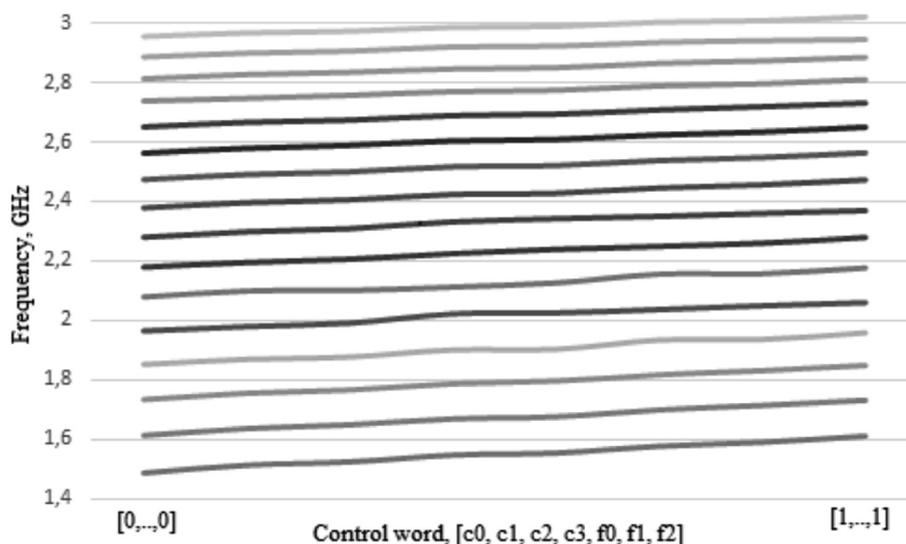


Fig. 4. Frequency of oscillation vs digital control word (Low band)

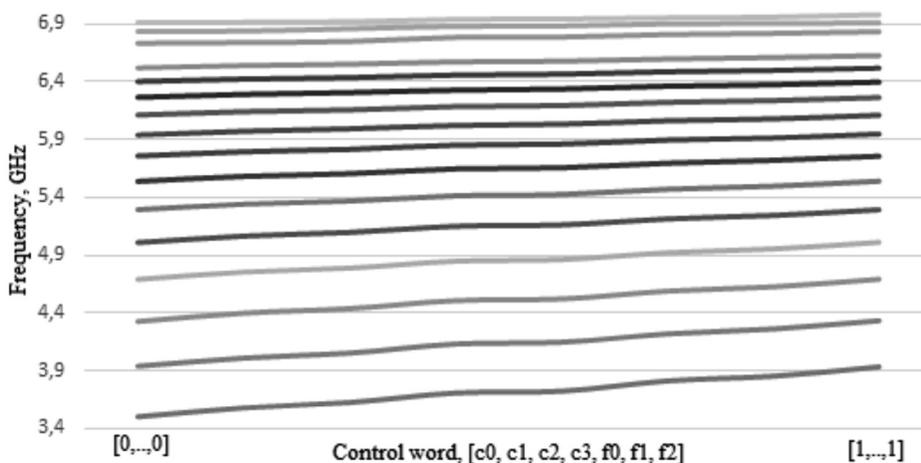


Fig. 5. Frequency of oscillation vs digital control word (High band)

Comparison of this study with the 65-nm technology

Ref.	Tech (nm)	Vdd (V)	Bit word	Frequency range	Power consumption	Phase noise (dBc/Hz)	Jitter (ps)	FOM (dBc/Hz)	Topology
This work	65	1	7 +1	Low band 1.48–3.02 GHz	0.359 mW @ 2.42 GHz	–84.4 @ 1 MHz	4.335	–156.5	Dual band DRO
				High band 3.5–6.98 GHz	1.86 mW @ 6.023 GHz				
[5]	65	1	5	47.8–538.7 MHz	0.142–0.205 mW	-	13.2 @ 64.4 MHz	-	Cascading cell base
[6]	65	1.8	14	5 GHz	2.16 mW	–149.1 @ 10 kHz	0.42	-	DRO
[7]	65	1.2	-	4.1–6.5 GHz	18 mW	–145 @ 1 MHz	-	–186.6	QDCO
[8]	65	1.2	4	1.67–2.45 GHz	7 mW	–78 @ 1 MHz	-	-	L based DCO
[9]	65	1	14	5.01 GHz	3.7 mW	–150 to –107 @ 1 MHz	0.45	–247.4	LC DCO
[10]	65	0.45	2	3.2 MHz	90 nW	–95 @ 1 MHz.	6.39	–150.6	DCLO
[11]	65	0.5	-	43.5–152 MHz	59 μ W	–103.4 @ 1 MHz	-	–155.7	Relaxation DCO

The low band and high band cover approximately 1.5 and 3.4 GHz frequency range respectively. The covered frequency range is enough to compensate the influence of the technological process and the temperature on the DCO tuning range.

The comparison of the proposed dual band DCO with the other study is summarized in Table.

Conclusion

The proposed DCO is designed in the 65-nm UMC CMOS technology. It achieves 0.35 mW and 1.86 mW power consumption at 2.4 and 5 GHz bands respectively. The expected jitter for both bands is not more than 4.33 ps and FOM is equal to 156.5 dBc/Hz. The phase noise is less than –84.4 dBc/Hz for both bands.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. Annema A.-J., Nauta B., van Langevelde R., Tuinhout H. Analog circuits in ultra-deep-submicron CMOS. *IEEE J. Solid-state Circuits*, 2005, Vol. 40, No. 1, Pp. 132–134.
2. Lewyn L., Ytterdal T., Wulff C., Martin K. Analog circuit design in nanoscale CMOS

technologies. *Proc. IEEE*, Oct. 2009, Vol. 97, No. 10, Pp. 1687–1714.

3. Pahlavan S., Ghaznavi Ghoushchi M.B. A 3.48 ps Jitter @ 1.45 GHz Fully Differential Dual band DCO with a new Reconfigurable Delay Cell. *Iranian Conference on Electrical Engineering*, 2016, 24th.

4. **Souri M., Ghaznavi Ghouschi M.B.** A 14.8 ps jitter low-power dual band all digital PLL with reconfigurable DCO and time-inter lined multiplexers. *Analog Integr. Circuits Signal Process*, 2015, Vol. 82, Pp. 381–392.
5. **Ching-Che Chung, Chiun-Yao Ko, Sung-En Shen** Built-in Self-Calibration Circuit for Monotonic Digitally Controlled Oscillator Design in 65-nm CMOS Technology. *IEEE Transactions on Circuits and Systems – II: Express Briefs*, 2011, Vol. 58, No. 3.
6. **Yuyu Chang, John Leete, Zhimin Zhou, Morteza Vadipour, Yin-Ting Chang, Hooman Darabi** A Differential Digitally Controlled Crystal Oscillator With a 14-Bit Tuning Resolution and Sine Wave Outputs for Cellular Applications. *IEEE Journal of Solid-state Circuits*, 2012, Vol. 47, No. 2.
7. **Shiyuan Zheng, Howard C. Luong** A 4.1-to-6.5 GHz Transformer-Coupled CMOS Quadrature Digitally-Controlled Oscillator with Quantization Noise Suppression. Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong.
8. **Ghorbel I., Haddad F., Barthélemy H., Rahajandraibe W., Loulou M., Mnif H.** Digitally controlled oscillator using active inductor based on CMOS inverters. *Electronics Letters*, 2014, Vol. 50, No. 22, Pp. 1572–1574.
9. **Ahmed Elkholy, Tejasvi Anand, Woo-Seok Choi, Amr Elshazly, Pavan Kumar Hanumolu** A 3.7 mW Low-Noise Wide-Bandwidth 4.5 GHz Digital Fractional-N PLL Using Time Amplifier-Based TDC. *IEEE Journal of Solid-state Circuits*, 2015, Vol. 50, No. 4.
10. **Dong-Woo Jee, Dennis Sylvester, David Blaauw, Jae-Yoon Sim.** Digitally Controlled Leakage-Based Oscillator and Fast Relocking MDLL for Ultra Low Power Sensor Platform. *IEEE Journal of Solid-state Circuits*, 2015, Vol. 50, No. 5.
11. **Yudong Zhang, Woogeun Rhee, Taeik Kim, Hojin Park, Zhihua Wang.** A 0.35–0.5-V 18–152 MHz Digitally Controlled Relaxation Oscillator With Adaptive Threshold Calibration in 65-nm CMOS. *IEEE Transactions on Circuits and Systems – II: Express Briefs*, 2015, Vol. 62, No. 8.

Received / Статья поступила в редакцию 14.08.2017

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

AKHMETOV Denis B.

АХМЕТОВ Денис Булатович

E-mail: akhmetov@spbstu.ru

AL-KARKHI Omar J.

E-mail: omarjalal81@gmail.com



Программное обеспечение вычислительных, телекоммуникационных и управляющих систем

DOI: 10.18721/JCSTCS.10306

УДК 519.682.1

ПОЛНАЯ ГОЛОВНАЯ ЛИНЕЙНАЯ РЕДУКЦИЯ

Д.А. Березун

Санкт-Петербургский государственный университет,
Санкт-Петербург, Российская Федерация

Головная линейная редукция (head linear reduction) представляет собой стратегию редукции лямбда-термов, производящую минимальное количество подстановок для достижения псевдоголовной нормальной формы (quasi-head normal form). Статья посвящена обобщению понятия головной линейной редукции до полной головной линейной редукции (complete head linear reduction), позволяющей полностью нормализовать лямбда-терм и определить новый подход к вычислениям — трассирующую нормализацию (traversal-based normalization). Оба подхода формализованы в виде систем переходов (transition system). В статье также показана корректность обеих стратегий редукций: головной линейной редукции относительно головной редукции — головная линейная редукция завершается в псевдоголовной нормальной форме терма тогда и только тогда, когда завершается головная, и полной головной линейной редукции относительно эффективной редуцирующей стратегии — головная линейная редукция завершается в нормальной форме терма тогда и только тогда, когда последняя существует.

Ключевые слова: лямбда-исчисление; редукция; линейная редукция; головная линейная редукция; полная головная линейная редукция; трассирующая нормализация.

Ссылка при цитировании: Березун Д.А. Полная головная линейная редукция // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2017. Т. 10. № 3. С. 59–82. DOI: 10.18721/JCSTCS.10306

COMPLETE HEAD LINEAR REDUCTION

D.A. Berezun

St. Petersburg State University, St. Petersburg, Russian Federation

In lambda calculus, head linear reduction is a reduction strategy which reaches a quasi-head normal form of terms in the minimum number of linear substitution steps. The paper is dedicated to the generalization of head linear reduction to a complete head linear reduction which yields normal forms when they exist. The formal presentation of both head linear reduction and complete head linear reduction via transition systems is provided. We also proved that both reduction strategies are correct: head linear reduction with respect to head reduction, i.e., that head linear reduction terminates in quasi-head normal form if and only if head reduction terminates, and we proved that

complete head linear reduction is an effective reduction strategy, i.e., it terminates if and only if the normal form exists.

Keywords: lambda-calculus; reduction strategy; linear reduction; head linear reduction; complete head linear reduction; traversal-based normalization.

Citation: Berezun D.A. Complete Head Linear Reduction. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems. 2017, Vol. 10, No. 3, Pp. 59–82. DOI: 10.18721/JCSTCS.10306

1 Введение

Головная линейная редукция играет особую роль в различных подходах к вычислениям, таких как игровая семантика (game semantics) [22–25], оптимальные редукции (optimal reductions) [16], геометрия взаимодействия (geometry of interaction) [21], сети доказательств (proof nets) [20] и др. Основанная на расширенном понятии редекса, именуемого *простым редексом* (prime redex), позволяющим явно линейризовать последовательность подстановок, головная линейная редукция не является редукцией лямбда-термов в её классическом понимании. Неформально говоря, головная линейная редукция вместо устранения редекса как такового и произведения подстановки в терм, запоминает редекс и производит линейные подстановки «по надобности», когда это необходимо, для достижения *псевдоголовной нормальной формы* (quasi-head normal form), т. е. такой формы терма, в которой головной переменной не соответствует ни один редекс. Головная нормальная форма получается из псевдоголовной нормальной формы путём подстановки всех простых редексов (см. раздел 3) во всех аргументах терма.

В статье приводится обобщение понятия головной линейной редукции до *полной головной линейной редукции* (complete head linear reduction), *приводящей терм в нормальную форму*. Полная головная линейная редукция позволяет определить новый подход к вычислениям — *трассирующую нормализацию* (traversal-based normalization) [6] — не использующую классические приемы, такие как контексты, замыкания и др.

В статье также показана корректность обеих стратегий редукций: приведено формальное доказательство того, что головная и полная головная линейные редукции завершаются в псевдоголовной нормальной и нормальной формах входного терма, со-

ответственно, тогда и только тогда, когда последние существуют. Иными словами, в общепринятой терминологии [5] полная головная линейная редукция является *эффективной редуцирующей стратегией* (effective reduction strategy).

2 Лямбда-исчисление

Лямбда-исчисление (lambda-calculus) впервые было предложено в 30-х гг. американским математиком Алонсо Чёрчем (Alonso Church) с целью формализации и анализа понятия вычислимости. Формальная система, предложенная Чёрчем, основана всего на трёх примитивах: переменной, абстракции, т. е. анонимной функции, и применении функции к аргументу. *Лямбда-выражением* (лямбда-термом, λ -термом или просто термом, Λ -term) называется выражение, удовлетворяющее грамматике $\Lambda^1 ::= V \mid \Lambda @^2 \Lambda \mid \lambda V$. Λ , где V — множество конечных строк, называемых *переменными*, над некоторым фиксированным алфавитом $\Sigma \setminus \{ , . , @ , \lambda \}$, выражение вида $\lambda x.e$ называется *абстракцией* по переменной x (λ -abstraction), а выражение вида $e_1 @ e_2$ — *применением* (application) терма e_1 к терму e_2 . Представленное в таком виде исчисление также называется *чистым* или *нетипизированным*, или *бестиповым* лямбда-исчислением.

Вхождение переменной в терм бывает двух видов: свободное и связанное.

¹ Мы также будем использовать скобки (“(”, “)”) как незначащие символы грамматики в тех ситуациях, когда из конкретного синтаксиса невозможно однозначное восстановление абстрактного.

² Символ @ иногда опускается: $\Lambda @ \Lambda \equiv \Lambda \Lambda \equiv \Lambda \Lambda$.

³ Мы будем считать, что приоритет оператора “@” выше чем у “.”, то есть $\lambda x.x @ y \equiv \lambda x.(x @ y)$, а ассоциативность у оператора “@” — левая, то есть $x @ y @ z = (x @ y) @ z$.

Связанными называются все вхождения переменных, по которым выше в дереве разбора были произведены абстракции, остальные вхождения переменных называются *свободными*. Так, например, в терме $\lambda y.x @ y @ (\lambda x.x)$ первое вхождение переменной x является свободным, второе — связанным, как и единственное вхождение переменной y .

Основной формой эквивалентности лямбда-термов является так называемая α -эквивалентность, утверждающая равенство термов, получающихся друг из друга переименованием связанных переменных. Так, например, термы $\lambda x.x$ и $\lambda y.y$ α -эквивалентны, поскольку первый получается из второго переименованием связанной переменной y в x .

Нередко термы в лямбда-исчислении рассматриваются с точностью до α -эквивалентности. Поэтому вместо терминов связанное и свободное вхождение переменной часто используют термины *связанная* и *свободная* переменная, соответ-

ственно, подразумевая, что одна и та же переменная не входит в терм свободно и связано одновременно. Более того, мы будем придерживаться *соглашения Барендрегта (Barendregt's convention), подразумевающего*, что все переменные имеют различные имена. Формально определения множеств свободных (FV) и связанных (BV) переменных приведены на рис. 1.

Основной аксиомой лямбда-исчисления является β -редукция. Интуитивно, β -редукция производит применение функции, т. е. лямбда-абстракции, к её аргументу путём замены всех вхождений переменной, связанной этой абстракцией, на тело аргумента:

$$(\lambda x.e_1) e_2 =_{\beta} e_1[x / e_2].$$

Такая замена называется *подстановкой* аргумента e_1 в терм e_2 , а само выражение вида $(\lambda x.e_1) e_2$ — *редексом*. Снабжённое аксиомой β -редукции λ -исчисление обладает свойством *полноты по Тьюрингу*, определяя тем самым одну из простейших моделей

Синтаксис
$\Lambda ::= V \mid \Lambda @ \Lambda \mid \lambda V.\Lambda$
$V ::= \{x, y, \dots\}$
Подстановка
$x[x/r] = r$
$y[x/r] = y, \text{ если } x \neq y$
$(e_1 e_2)[x/r] = (e_1[x/r])(e_2[x/r])$
$(\lambda x.e)[x/r] = \lambda x.e$
$(\lambda y.e)[x/r] = \lambda y.e[x/r], \text{ если } x \neq y \text{ и } y \notin FV(e)$
β -редукция
$(\lambda x.e_1) e_2 =_{\beta} e_1[x/e_2]$
Множество свободных переменных
$FV(x) = \{x\}$
$FV(\lambda x.e) = FV(e) \setminus \{x\}$
$FV(e_1 e_2) = FV(e_1) \cup FV(e_2)$
η -конверсия
$f \sim_{\eta} \lambda x.f x, \text{ если } x \notin FV(f)$

Синтаксис в нотации де Брауна

$$\Lambda^d ::= \mathbb{N} \mid \Lambda^d \Lambda^d \mid \lambda \Lambda^d$$

Сдвиг в Λ^d

$$\uparrow_c^d(k) = \begin{cases} k, & k < c \\ k + d, & k \geq c \end{cases}$$

Подстановка в Λ^d

$$k[j/r] = \begin{cases} r, & k = j \\ k, & \text{иначе} \end{cases}$$

$$\uparrow_c^d(\lambda e) = \lambda \uparrow_{c+1}^d(e) \quad (\lambda e)[j/r] = \lambda e[j + 1 / \uparrow_0^1 r]$$

$$\uparrow_c^d(e_1 e_2) = \uparrow_c^d(e_1) \uparrow_c^d(e_2) \quad (e_1 e_2)[j/r] = e_1[j/r] e_2[j/r]$$

β -редукция в Λ^d

$$(\lambda e_1) e_2 \rightarrow \uparrow_0^{-1}(e_1[0 / \uparrow_0^1(e_2)])$$

Множество связанных переменных

$$BV(x) = \emptyset$$

$$BV(\lambda x.e) = BV(e) \cup \{x\}$$

$$BV(e_1 e_2) = BV(e_1) \cup BV(e_2)$$

α -эквивалентность

$$\lambda x.e \sim_{\alpha} \lambda y.e[x/y], \text{ если } y \notin FV(e)$$

Рис. 1. Чистое лямбда-исчисление

вычислений. Термы, получаемые друг из друга по правилу β -редукции, называются β -эквивалентными. Запись $s \rightarrow_{\beta} t$ означает, что терм t получается из терма s за один шаг β -редукции, а запись $s \rightarrow_{\beta}^* t$ означает, что терм t получается из терма s за ноль или более шагов β -редукции.

Заметим, что подстановка не является тривиальной операцией. Так, например, рассмотрим терм $(\lambda x.x y) x$. Согласно правилу α -эквивалентности, он эквивалентен терму $(\lambda z.z y) x$, который, в свою очередь, β -эквивалентен терму $\lambda z.z x$. Если произвести подстановку $[x / y]$ непосредственно в терме $\lambda x.x y$, то результатом будет терм $\lambda x.x x$, который уже не эквивалентен терму $\lambda z.z x$. Таким образом, для подстановки важно, чтобы переменные, встречающиеся свободно в подставляемом терме, не связывались при подстановке. Такая подстановка называется *свободной от связывания (capture-avoiding substitution)*, далее — просто подстановка. Формальные правила подстановки приведены на рис. 1.

В начале 70-х гг. голландским математиком Николасом де Брауном⁴ (Nicolaas Govert de Bruijn) была предложена нотация, получившая название *представление де Брауна* [2], позволяющая полностью избавиться от имён в лямбда-термах. В представлении де Брауна терм полностью лишается имён переменных: вхождения переменных заменяются на натуральные числа, называемые *индексами де Брауна (de Bruijn indexes)*, а абстракции становятся анонимными. По сути, термы в представлении де Брауна представляют собой классы эквивалентности термов, факторизованных по правилу α -эквивалентности. *Индекс де Брауна* суть натуральное число, представляющее собой вхождение переменной, равное количеству абстракций между этим вхождением и абстракцией, связывающей саму переменную, включительно. Так, например, терм $\lambda x.\lambda y.\lambda z.x z (y z)$, также известный как S

комбинатор, в нотации де Брауна имеет следующий вид: $\lambda \lambda \lambda 3 1 (2 1)$. Заметим, что в представлении де Брауна при осуществлении подстановки может потребоваться пересчёт индексов (см. рис. 1).

Ещё одно преобразование иногда считается стандартным для лямбда-исчисления — η -конверсия (эта-конверсия, η -coersion), интуитивно, утверждающее, что функция и абстракция этой функции, применённой к переменной, по которой осуществлена абстракция, суть одно и то же (см. рис. 1). Преобразование функции f к виду $\lambda x.fx$ получило название η -расширения (η -expansion), а преобразование, ему обратное, — η -редукции (η -reduction).

2.1 Стратегии вычислений

Говорят, что терм находится в *нормальной форме*, если он не содержит редексов, т.е. к нему неприменима аксиома β -редукции, а два терма называются *равными*, с точностью до α -конверсии, если они имеют одну и ту же нормальную форму. Вычисление в лямбда-исчислении есть вычисление нормальной формы терма — нормализация. Конечно, ввиду Тьюринг-полноты λ -исчисления не все термы имеют нормальную форму, например, вычисление терма, также известного как терм омега (ω , omega), $\omega = \sigma\sigma = (\lambda x.x x) (\lambda x.x x)$ расходится. Важным свойством в контексте нормализации является свойство Чёрча–Россера (Church–Rosser), также известное как свойство ромба, утверждающее, что если два терма s и t равны, то существует такой терм m , что $s \rightarrow_{\beta}^* m$ и $t \rightarrow_{\beta}^* m$. Прямым следствием свойства ромба является единственность нормальной формы терма, если таковая существует. Тем не менее, в аксиоме β -редукции ничего не сказано про порядок, в котором должна происходить β -редукция терма. Далее в этой части мы рассмотрим существующие *порядки* (стратегии, reduction strategies) *редукций* и их свойства.

Разнообразие стратегий вычислений в лямбда-исчислении обусловлено, в том числе, тем, что разные языки программирования изобретались для различных целей и, соответственно, имеют разные свойства и реализации. Несмотря на то, что не все

⁴ В русскоязычной литературе также встречаются следующие варианты перевода фамилии «de Bruijn»: «де Брюйн», «де Брейн», «де Бройн».

Редукция аргументов	Редукция под абстракцией	
	Да	Нет
Да	(Сильная) нормальная форма $S ::= \lambda x.N \mid x N_1 \dots N_n$	Слабая нормальная форма $W ::= \lambda x.e \mid x W_1 \dots W_n$
Нет	Головная нормальная форма $H ::= \lambda x.H \mid x e_1 \dots e_n$	Слабая головная нормальная форма $E ::= \lambda x.e \mid x e_1 \dots e_n$

где $\forall i \in \mathbb{N}$, e_i — произвольный лямбда-терм

Рис. 2. Нормальные формы⁶

порядки редукций приводят терм к нормальной форме, они нередко завершаются в некоторой «нормальной»⁵ для этого порядка форме, которая с точки зрения конкретного порядка редукций считается *значением*, т. е. термом, который не редуцируется дальше, даже если содержит редексы. Основными свойствами стратегий редукций являются следующие два: произведение редукций под абстракцией и редукция аргументов. В таблице на рис. 2 приведены «нормальные» формы, получаемые при всех комбинациях этих свойств, описанные в терминах контекстно-свободных грамматик⁶.

2.1.1 Слабые порядки редукций

Слабые порядки редукций характеризуются тем, что они считают абстракцию значением и, соответственно, не производят редукцию под абстракцией.

Вызов по имени (call-by-name). Интуитивно, стратегия вычисления по имени сначала вычисляет функцию до значения, лямбда-абстракции, после чего редуцирует её применение к аргументу путём постановки тела аргумента вместо каждого вхождения пере-

менной, по которой эта абстракция произведена, что может привести к повторным вычислениям аргументов, если вхождений переменной, по которой производилась абстракция функции, несколько. Формально, на каждом шаге раскрывается самый левый, самый внешний редекс, который не находится под лямбда-абстракцией, и, таким образом, вызов по имени завершается в слабой головной нормальной форме.

Вызов по значению (call-by-value) в отличие от вызова по имени сначала производит редукцию аргументов и лишь потом применения функции к ним. На каждом шаге раскрывается самый левый, самый внутренний редекс, который не находится под лямбда-абстракцией, и, таким образом, вызов по значению завершается в слабой нормальной форме.

Вызов по необходимости (call-by-need) является вызовом по имени, в который добавлена *мемоизация*, т. е. при вычислении аргумента его значение сохраняется и не перевычисляется при последующем повторном использовании аргумента. В *чистых функциональных языках*, где нет побочных эффектов, результат вызова по имени и вызова по необходимости будет одинаковым, а значит, вызов по необходимости завершается в слабой головной нормальной форме. Одним из наиболее популярных языков, использующих вызов по необходимости, является язык функционального программирования Haskell.

2.1.2 Сильные порядки редукций

Нормальный порядок (applicative order) отличается от стратегии вызова по имени тем, что он производит *сильную* редукцию, т. е. редукцию под абстракцией, а также

⁵ Данная форма не является нормальной формой в том смысле, что терм, находящийся в ней, может содержать редексы. Для того, чтобы такая терминология не вводила в заблуждение, нормальную форму часто называют *сильной* нормальной формой, а остальные «нормальные» формы имеют некоторый дополнительный префикс, описывающий «какая именно это нормальная форма», например, головная нормальная форма (см. главу 2.2.3 и рис. 2).

⁶ Таблица заимствована из статьи Питера Сестофта (Peter Sestoft) [7].

редукцию аргументов. Первым на каждом шаге раскрывается самый левый, самый внешний редекс. В отличие от других порядков редукций, нормальный порядок редукций является *нормализующим* (normalizing), т. е. завершается в сильной нормальной форме терма тогда и только тогда, когда последняя существует. С другой стороны, как и вызов по имени, нормальный порядок редукций может проделывать одну и ту же работу несколько раз. Например, $(\lambda x.xx)((\lambda y.y)a) \rightarrow ((\lambda y.y)a)((\lambda y.y)a) \rightarrow a((\lambda y.y)a) \rightarrow aa$ дважды редуцирует редекс $(\lambda y.y)a$.

Аппликативный порядок (applicative order) В то время как нормальный порядок является сильной версией вызова по имени, аппликативный порядок является сильной версией вызова по значению. Иными словами, на каждом шаге первым раскрывается самый левый, самый глубокий редекс. Например, $(\lambda x.xx)((\lambda y.y)a) \rightarrow (\lambda x.xx)a \rightarrow aa$. Если аппликативный порядок завершается, то приводит терм в сильную нормальную форму. Тем не менее, в отличие от нормального порядка редукций, аппликативный порядок может не завершаться, даже если терм имеет нормальную форму. Нередко это происходит тогда, когда какой-нибудь аргумент функции не имеет нормальной формы и при этом не используется в её теле. Например, терм $(\lambda x.\lambda y.x)a((\lambda x.xx)(\lambda x.xx))$ имеет нормальную форму a , в то время как его аргумент $((\lambda x.xx)(\lambda x.xx))$, уже известный нам терм ω нормальной формы не имеет, и его вычисление расходится. Более того, аппликативный порядок редукций не способен редуцировать терм, являющийся применением рекурсивной функции к аргументу, даже если комбинатор рекурсии, используемый для определения рекурсивной функции, является специальным комбинатором для вызова по значению.

Головная редукция (head reduction, leftmost head reduction) производит редукцию только головных редексов. Редекс называется *головным*, если его предками, в смысле абстрактного синтаксического дерева терма, являются лишь абстракции. Например, редекс $(\lambda y.e_1)e_2$ является головным в терме $\lambda x_1 \dots \lambda x_n.(\lambda y.e_1)e_2 \dots e_m$. Завершается голов-

ная редукция в головной нормальной форме. Заметим, что повторное применение головной редукции к аргументам полученной головной нормальной формы нормализует терм. Более того, такая стратегия редукций является нормализующей.

Существуют и другие порядки редукций, нередко являющиеся лишь вариациями и разнообразными комбинациями описанных выше порядков редукций, такие как *гибридный* (hybrid) аппликативный порядок, гибридный нормальный порядок или *спиновая головная редукция* (spine head reduction) и др. Мы опустим их в данной статье, обратив читателя к [5, 7].

2.2 Головная линейная редукция и трассирующая нормализация

Как уже отмечалось ранее, лямбда-исчисление в первую очередь представляет собой модель вычислений, и, соответственно, одним из основных вопросов, стоящих перед ней, является вопрос о *сложности* (complexity) вычислений. Иными словами, положим, дан некоторый терм, и зафиксирована стратегия вычислений. Какова сложность вычислений – нормализации – данного терма. Простейшая мера – количество шагов β -редукции, необходимых для нормализации терма. К сожалению, такой подход не соответствует сложности вычислений на реальных вычислителях, поскольку сложность операции подстановки находится в непосредственной нелинейной зависимости от терма.

Работы [13–18] представляют различные подходы к описанию сложности вычислений в лямбда-исчислении и определению оптимальных стратегий редукций, основанных на той или иной мере сложности вычислений. Одним из таких подходов является *головная линейная редукция* (head linear reduction) [3, 4], на каждом шаге вместо обычных подстановок производящая так называемые *линейные подстановки* (linear substitution) – подстановки только одного вхождения переменной в терм. В этой главе мы рассмотрим классические определения головной линейной редукции, трассирующей нормализации и связь между ними.

Терм T	Условие	Список головных абстракций $\lambda_h(T)$	Список простых редексов $pr(T)$
x		$[\]$	$[\]$
UV	$\lambda_h(U) = [\]$	$[\]$	$pr(U)$
UV	$\lambda_h(U) = \lambda x : l$	l	$(\lambda x, V) : pr(U)$
$\lambda x.U$		$\lambda x : \lambda_h(U)$	$pr(U)$

Рис. 3. Определение списков головных абстракций и простых редексов

2.2.1 Головная линейная редукция: классическое определение

Для формального определения понятия головной линейной редукции нам понадобится ввести некоторое количество дополнительных определений. *Хребетными* или *спинальными подтермами* (spine sub-terms) терма T называются сам терм T , а также все спинальные подтермы терма U , если $T = UV$ или $T = \lambda x.U$ соответственно. Заметим, что любой лямбда-терм имеет ровно один спинальный подтерм-переменную. Такое вхождение переменной, так же как и сам подтерм, называется *головным* (head occurrence, hoc). Ещё два понятия: *список головных абстракций* (head λ -list, $\lambda_h(T)$) и *простые редексы* (prime redexes) — определяются по индукции по структуре терма T . Простой редекс представляет собой пару $(\lambda x, N)$, первый и второй элементы которой называются *абстракцией* и *аргументом простого редекса* соответственно. Индуктивное определение списков простых редексов и головных абстракций приведены

на рис. 3, где $pr(T)$ обозначает список простых редексов терма T .

Редексом головного вхождения переменной (hoc redex) терма T называется простой редекс $(\lambda x, V)$, где x является головной переменной, если таковой редекс существует.

Интуитивно, головная линейная редукция линеаризует последовательность подстановок путём замены на каждом шаге лишь одного вхождения переменной (головного), оставляя сам головной редекс нетронутым. Если же он не определён, головная линейная редукция завершается в так называемой *псевдоголовной нормальной форме* (quasi-head-normal form, qhn).

Пусть $r = (\lambda x, \dots)$ и $s = (\lambda y, \dots)$ — два простых редекса терма T . Говорят, что простой редекс r *содержит* простой редекс s , если λy является вершиной поддерева, начинающегося в вершине λx . Более того, простые редексы r и s называются *последовательными*, если r содержит s , и не существует такого простого редекса t , который содержится в r , но не содержится в s .

Пример. Пусть $T = \lambda s.(\lambda x.(\lambda y.(\lambda w.w @ s) @ y) @ x) @ (\lambda z.z)$, тогда

$$\begin{cases} \lambda_h(T) = [\lambda s] \\ pr(T) = [(\lambda x, (\lambda z.z)), (\lambda y, x), (\lambda z, y)] \iff [r, s, t] \end{cases}$$

Пары (r, s) и (s, t) образуют последовательные простые редексы.

Теорема 1. Пусть N — произвольный терм, тогда если терм M является результатом применения головной линейной редукции к терму N , то:

1. $M \equiv_{\beta} N$;

2. Головная линейная редукция завершается тогда и только тогда, когда завершается головная редукция.

Первое утверждение может быть показано индукцией по числу шагов головной линейной редукции. Что же касается вто-

рого утверждения, то его необходимость устанавливается в силу того, что головная редукция терма N завершится за число шагов, равное количеству простых редексов терма M . Действительно, никакая β -редукция терма M не способна произвести подстановку головного вхождения переменной, а значит, и не способна создать новый простой редекс. Таким образом, терм имеет головную нормальную норму, что в свою очередь гарантирует завершаемость [5] го-

Синтаксис по Карри	Типизация по Карри	Типизация по Чёрчу
$\Lambda^{Cu} ::= \Lambda$		
Синтаксис по Чёрчу	$\frac{x : \tau \in \Gamma}{\Gamma \vdash_{Cu} x : \tau}$	$\frac{x : \tau \in \Gamma}{\Gamma \vdash_{Ch} x : \tau}$
$\Lambda^{Ch} ::= V \mid \Lambda^{Ch} \Lambda^{Ch}$		
$\quad \mid \lambda V^{\tau} . \Lambda^{Ch}$	$\frac{\Gamma, x : \tau \vdash_{Cu} e : \sigma}{\Gamma \vdash_{Cu} \lambda x . e : \tau \rightarrow \sigma}$	$\frac{\Gamma, x : \tau \vdash_{Ch} e : \sigma}{\Gamma \vdash_{Ch} \lambda x^{\tau} . e : \tau \rightarrow \sigma}$
Типы		
$\tau ::= \iota \mid \tau \rightarrow \tau$	$\frac{\Gamma \vdash_{Cu} e_1 : \tau_1 \rightarrow \tau_2 \quad \Gamma \vdash_{Cu} e_2 : \tau_2}{\Gamma \vdash_{Cu} e_1 e_2 : \tau_2}$	$\frac{\Gamma \vdash_{Ch} e_1 : \tau_1 \rightarrow \tau_2 \quad \Gamma \vdash_{Ch} e_2 : \tau_2}{\Gamma \vdash_{Ch} e_1 e_2 : \tau_2}$
Контексты		
$\Gamma ::= \emptyset \mid \Gamma, V : \tau$		

Рис. 4. Простое типизированное лямбда-исчисление

ловной редукции. Тем не менее доказать достаточность утверждения 2 не так просто. Для этого мы введём формальное определение головной линейной редукции в виде системы переходов, с помощью которой формально и докажем утверждение теоремы (см. главу 3).

2.2.2 Простое типизированное лямбда-исчисление

Простое типизированное лямбда-исчисление (simply-typed lambda-calculus, STLC) имеет схожий с бестиповым синтаксис, тем не менее, не все термы бестипового исчисления являются *допустимыми* (valid) термами простого типизированного. Типы в простом типизированном лямбда-исчислении описываются следующей грамматикой: $\tau ::= \iota \mid \tau \rightarrow \tau$, где ι представляет собой некоторый *базовый тип* (ground type), а $\tau_1 \rightarrow \tau_2$ — *функциональный тип* (стрелочный тип, arrow type), ассоциативность у оператора *стрелка* \rightarrow правая. Принято выделять два похода, два стиля типизации: стиль *Карри* (типизация по Карри, \a la Curry) и стиль *Чёрча* (типизация по Чёрчу, \a la Church). Эти два подхода являются принципиально разными: при типизации по Карри сначала задаётся грамматика термов — синтаксис, затем определяется их поведение — семантика, и наконец вводится система типов — типизация, отвергающая термы, обладающие нежелательным поведением, в то время как в стиле Чёрча типизация предшествует семантике⁷. Иными словами, при типизации по Чёрчу семантикой — смыслом — наделены исключительно *пра-*

ильно типизированные (well-formed, well-typed) термы, а типизация по Карри даёт возможность рассуждать о поведении лямбда-термов вне зависимости от того, являются ли они правильно типизированными или нет. Синтаксис и типизации по Чёрчу и по Карри простого типизированного лямбда-исчисления в виде правил вывода (inference rules) приведены на рис. 4. Суждение вида $\Gamma \vdash \Lambda : \tau$ называется *терм в контексте* (term-in-context), где *контекст типизации* Γ (typing context) является множеством текущих предположений о типах термов.

Одним из основных отличий двух систем типизации является свойство единственности типа: при типизации по Чёрчу терм имеет не более одного типа, в то время как при типизации по Карри терм может иметь один или несколько типов, или быть не типизируемым вовсе. Например, терм $\lambda x : \tau . x$ (при типизации по Чёрчу) имеет един-

⁷ Иногда под типизацией по Чёрчу понимают явно типизированные системы (типы переменных явно указываются в синтаксисе языка), а под типизацией по Карри — неявно типизированные (типы присваиваются в процессе компиляции или интерпретации программы). Путаница возникла ввиду того, что сам Чёрч описывал своё исчисление в явном стиле, в то время как Карри использовал неявный стиль. Несмотря на то, что такой взгляд на стили типизации является исторически сложившимся: явно типизированные системы часто описывались и описываются в стиле Чёрча, а неявно типизированные — в стиле Карри, — он всё же является ошибочным.

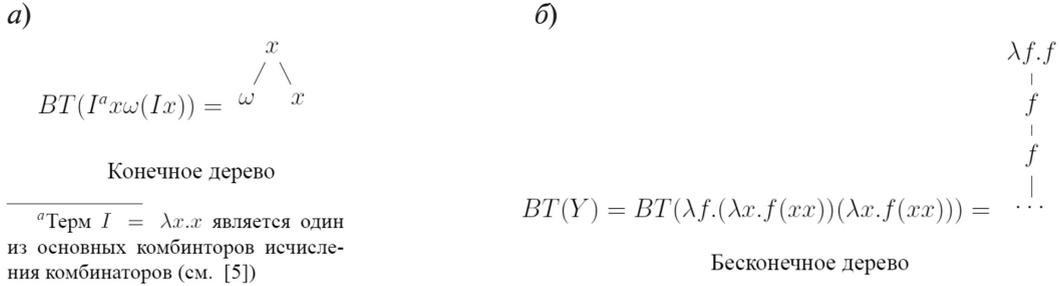


Рис. 5. Пример: деревья Бёма для термов, не имеющих нормальной формы

ственный тип $\tau \rightarrow \tau$, где τ – некоторый конкретный тип, указанный в синтаксисе, например, ι или $(\iota \rightarrow \iota \rightarrow \iota) \rightarrow \iota \rightarrow \iota \rightarrow \iota$ и т. д., а аналогичному ему терму $\lambda x.x$ в представлении Карри соответствует всё множество типов вида $\alpha \rightarrow \alpha$.

Как уже упоминалось, не все термы бестипового исчисления являются термами простого типизированного. Более того, в отличие от бестипового лямбда-исчисления, простое типизированное является *строго нормализуемым* (strongly normalised), т. е. каждый допустимый терм имеет нормальную форму, а процедура его нормализации всегда завершается. Так, например, терм бестипового лямбда-исчисления ω не имеет нормальной формы и не является допустимым термом простого типизированного. Исчерпывающее описание простого типизированного лямбда-исчисления и его свойств, а также более богатых систем типов читатель может найти в [5, 10, 11].

2.2.3 Головная нормальная форма и дерево Бёма

Произвольный лямбда-терм M можно записать в следующей форме⁸:

$$BT(M) := \begin{cases} \omega & , \text{ если } M \text{ не имеет hnf} \\ \begin{array}{c} \lambda \vec{x}.y \\ / \quad | \quad \backslash \\ BT(V_1) \quad \dots \quad BT(V_n) \end{array} & , \text{ если } \text{phnf}(M) = \lambda \vec{x}.yV_1 \dots V_n \end{cases}$$

Как уже отмечалось ранее, повторное применение головной редукции к аргумен-

⁸ Утверждение может быть легко доказано индукцией по структуре терма M .

$$M = \lambda \vec{v}.UV_1 \dots V_n, \text{ где } U = \begin{cases} y \\ (\lambda y.P)Q \end{cases} \quad (1)$$

В первом случае говорят, что терм M находится в *головной нормальной форме* (head normal form), во втором случае $(\lambda y.P)Q$ образует головной редекс. Заметим, что по определению терм находится в головной нормальной форме тогда и только тогда, когда он не содержит головных редексов.

Важным следствием этого определения является то, что головная нормальная форма не обладает свойством единственности: один и тот же терм может иметь более одной головной нормальной формы. Например, головной нормальной формой терма $y((\lambda x.x)z)$ является как терм сам по себе, так и его нормальная форма yz . Среди множества головных нормальных форм терма выделяют *основную* головную нормальную форму (principal head normal form, phnf), получаемую посредством головной редукции терма, записанного в виде (1), если редукция завершается.

Деревом Бёма (Böhm tree) $BT(M)$ терма M называется дерево, построенное по следующим правилам:

там головной нормальной формы нормализует терм. Таким образом, если терм имеет нормальную форму, то дерево Бёма, представляющее этот терм, вычислимо и конечно, и наоборот [5, 27, 28]. Также отметим,

что если терм нормальной формы не имеет, то дерево Бёма, ему соответствующее, либо конечно, а процедура построения его следующего уровня расходится (см. рис. 5 а), либо является бесконечным (см. рис. 5 б).

2.2.4 η-длинная форма терма

η-длинной форма (η-long form) терма получается путём его полного η-расширения и заменой бинарного оператора применения на оператор *длинного применения* @^l (long application). Заметим, что каждая переменная, непосредственным предком которой не является абстракция, также подлежит η-расширению, путём введения анонимной (dummy) абстракции ($x \mapsto \lambda.x$). Например, η-длинной формой терма $\lambda x^{1 \rightarrow 1}.x$ является терм $\lambda x^{1 \rightarrow 1}.a^1 b^1.x(\lambda.b)(\lambda.a)$. Заметим, что понятие η-длинной форм имеет смысл только в типизированном исчислении,

$$\Lambda_{odd}^{lf} ::= \lambda x_1^{\tau_1} \dots x_p^{\tau_p}. \Lambda_{even}^{lf}$$

$$\Lambda_{even}^{lf} ::= x \Lambda_{1odd}^{lf} \dots \Lambda_{nodd}^{lf} \mid \Lambda_{0odd}^{lf} @^l \Lambda_{1odd}^{lf} \dots \Lambda_{modd}^{lf}, \text{ где } m \in \mathbb{N}, n, p \in \mathbb{N}_0$$

Иными словами, нечётные уровни в абстрактном синтаксическом дереве представляют собой абстракции по произвольному числу аргументов, а чётные — применением переменной к нулю или более аргументам, либо же применение другого η-длинного терма к одному или более аргументам, получившее название оператора длинного применения @^l.

2.2.5 Трассирующая нормализация

В 70-х гг. XX в. Гордон Плоткин [8] сформулировал проблему построения *полностью абстрактной модели вычислений* (fully abstract model of computations) для языка программирования вычислимых функций PCF (Programming Computable Functions), заключающуюся в построении абстрактной модели языка, являющейся одновременно *полной* (complete) — каждый терм языка представим в его модели — и *согласованной* (sound) — каждый элемент абстрактной модели имеет прообраз в языке. Впервые предложенная Плоткиным проблема была решена в начале 90-х гг. независимо двумя группами исследователей (Мартинот Хуландом и Люком Онгом [23], и Самсоном Абрамски и Гаем МакКаскером [22])

т. к. в нём η-расширение ограничено хотя бы размером типа терма, в то время как в бестиповом случае η-расширять терм можно до бесконечности.

β-нормальной η-длинной является форма терма, не содержащая β-редексов, но являющаяся η-длинной. Например, β-нормальной η-длинной формой терма $(\lambda x^{1 \rightarrow 1}.x) y$ является терм $\lambda a^1. y(\lambda.a)$, которой произведением полной η-редукции, которая, разумеется, всегда завершается, η-эквивалентен терму $y^{1 \rightarrow 1}$, являющемуся, в свою очередь, нормальной формой исходного терма.

Заметим, что абстрактный синтаксис, используемый в определении η-длинной формы терма отличается от стандартного абстрактного синтаксиса λ-исчисления, введённого в начале главы 2, а именно, он описывается следующей грамматикой:

посредством *игровой семантики* (game semantics) программ, являющейся одним из способов задания формальной семантики языков программирования. Игровая семантика рассматривает вычисления, как *игру* (game) между *пропонентом(-ами)* (proponent, player) и *оппонентом(-ами)* (opponent) — программой и её окружением, а семантика — смысл — программы — *стратегия* (strategy), которой игрок должен придерживаться. Подробнее с игровой семантикой читатель может ознакомиться в работах [22–26].

В 2015 г. Люком Онгом было замечено [1], что из игровой семантики программ для простого типизированного лямбда-исчисления естественным образом вытекает нестандартная процедура его нормализации. Отличительной особенностью данной процедуры является то, что вместо изменения терма посредством β-редукции, используя стандартные приёмы, такие как окружения или замыкания⁹, *трассирующая нормализация* (traversal-based normalization, normalization by traversals) оставляет входной терм нетронутым, производя его нормализацию путём обхода абстрактного синтаксического дерева терма, запоминая историю этого



обхода. Оригинальная процедура нормализации, предложенная Онгом, требует преобразования термов в η -длинную форму, завершаясь в β -нормальной η -длинной форме.

Интуитивно, процедура трассирующей нормализации, предложенная Онгом, производит обход абстрактного синтаксического дерева терма в глубину по самому левому пути, запоминая историю обхода, называемую *трассой* (traversal). Достигая вхождения связанной переменной, процедура нормализации «перепрыгивает» на поддереву, соответствующее динамическому аргументу, с которым связана абстракция этой переменной. Заметим, что такой аргумент всегда найдётся, в силу того, что терм находится в η -длинной форме, а значит, все абстракции применены к некоторым аргументам. При достижении вхождения свободной переменной, процедура нормализации «разделяется», т. е. продолжается независимыми обходами каждого из аргументов этой переменной, порождая тем самым множество трасс, каждая из которых впоследствии определит уникальный путь от вершины до некоторого листа в абстрактном дереве η -длинной β -нормальной

формы входного терма. Процедура обхода и построения трасс является синтаксически управляемой и детерминированной. Стоит также отметить, что предложенная процедура нормализации в некотором смысле¹⁰ соответствует головной линейной редукции термов.

2.3 Системы переходов

Одним из подходов к изучению поведения дискретных систем являются *системы переходов* (transition system, [9]) – суть четвёрка (S, I, F, \rightarrow) , где S – множество состояний системы, $I \subseteq S$ – множество исходных состояний системы, $F \subseteq S$ – множество конечных состояний системы, $\rightarrow \subseteq S \times S$ – отношение, означающее дискретный переход системы из одного состояния в другое, обозначаемое $(s_1, s_2) \in \rightarrow$ или, для удобства чтения, $s_1 \rightarrow s_2$, где $s_1, s_2 \in S$.

Система переходов называется *детерминированной*, если текущее состояние системы однозначно определяет последующее, в противном случае система называется *недетерминированной*.

Например, алгоритм Евклида может быть записан следующей системой переходов:

$$\begin{aligned} S &= \mathbb{N} \times \mathbb{N} \\ I &= S \\ F &= \{(n, n) \mid n \in \mathbb{N}\} \\ \rightarrow &= \{((m, n), (m - n, n)) \mid m > n\} \cup \{((m, n), (m, n - m)) \mid m < n\} \end{aligned}$$

3 Головная линейная редукция (HLR)

В этой главе мы введём формальное определение головной линейной редукции в виде системы переходов и докажем её согласованность с головной редукцией.

3.1 HLR как система переходов

Определение. Система переходов для головной линейной редукции:

⁹ *Замыканием* (closure) называется функция первого порядка, определяющая значения переменных определённых вне тела функции, в случае лямбда-исчисления – свободных переменных. Использование замыканий является одной из стандартных способов реализации функциональных языков программирования.

1. Состоянием является тройка $\langle A[\underline{B}]; \Gamma; \Delta \rangle$, где

- $A[\underline{B}]$ – λ -терм, в котором вершина, являющаяся корнем поддерева B , выделена¹¹.

¹⁰ Внимательный читатель мог заметить, что головная линейная редукция завершается в псевдоголовой нормальной форме, в то время как трассирующая нормализация в стиле Онга – в β -нормальной η -длинной форме. Тем не менее, с точностью до вхождения свободной переменной, имеющей не пустой список аргументов, и η -эквивалентности, если входной терм находится в η -длинной форме, эти формы равны.

¹¹ Выделение является синтаксической пометкой вершины абстрактного синтаксического дерева терма и обозначается *подчёркиванием* в примерах и правилах.

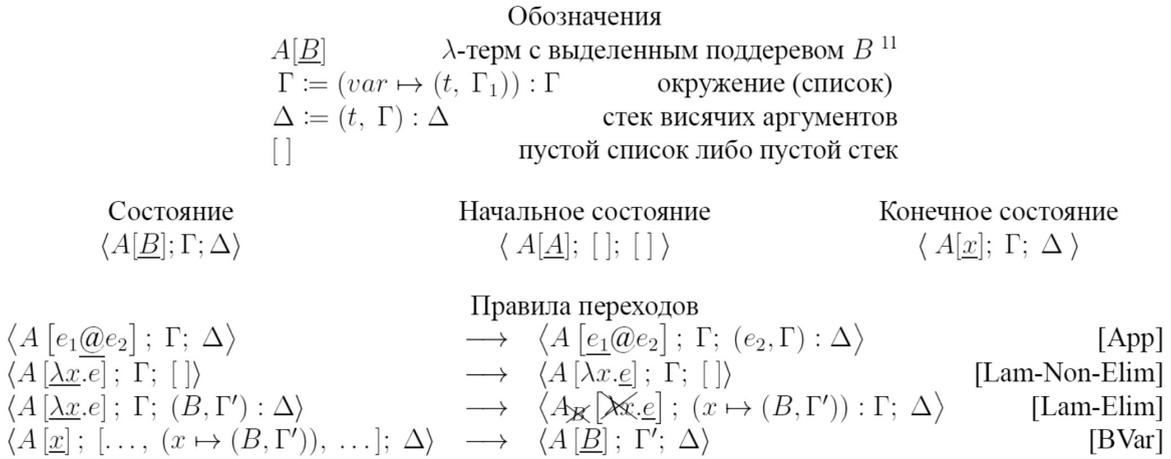


Рис. 6. Система переходов для головной линейной редукции

• $\Gamma := (var \mapsto (t, \Gamma_1)) : \Gamma$ — окружение — список связываний переменных, где var — переменная, t — λ-терм, а Γ_1 — сохранённое окружение, соответствующее терму t .

• $\Delta := (t, \Gamma) : \Delta$ — стек висячих аргументов — пар вида (t, Γ) , где t является λ-термом, а Γ — соответствующим ему окружением.

2. Начальным является состояние $\langle A[A]; []; [] \rangle$, где $[\]$ — пустые окружение и стек висячих аргументов, соответственно, а $A[A]$ — входной терм с выделенным корнем.

3. Конечным является состояние вида $\langle A[x]; \Gamma; \Delta \rangle$, где $A[x]$ — исходный λ-терм с выделенной переменной x , $x \notin \Gamma$.

4. Правила переходов приведены на рис. 6.

• В случае применения (правило [App]), выделенным становится его левый аргумент, а в стек висячих аргументов помещается аргумент с текущим окружением. В дальнейшем этот аргумент может использоваться для формирования простого редекса.

• Если выделенной вершиной является абстракция, то либо, если стек висячих аргументов пуст, и переменная абстракции не связывается ни с каким аргументом (правило [Lam-Non-Elim]), либо, если стек висячих аргументов не пуст, формирует простой редекс с его вершиной (правило [Lam-Elim]). В этом случае абстракция и соответствующие применение и его аргу-

мент вычёркиваются¹² из текущего дерева, а простой редекс сохраняется в текущем окружении. Очевидно, что правила напрямую согласованы с определением простого редекса.

• Если же выделенной является переменная, то, если она является свободной, система переходов достигла своего конечного состояния, ежели переменная является связанной, то либо существует простой редекс, аргумент которого может быть подставлен вместо вхождения этой головной переменной (правило [BVar]), либо такого редекса нет, и опять же система переходов достигла конечного состояния.

Заметим, что система переходов является детерминированной и синтаксически управляемой, а выбор правила зависит лишь от первого элемента состояния (формально, выбор между правилами [Lam-*] зависит от текущего состояния стека висячих аргументов, но он также напрямую зависит от входного терма). Единственным случаем, когда система переходов может достичь своего конечного состояния, является случай выделенной переменной, которая либо свободна, либо не связана никаким про-

¹² Вычёркивание является синтаксической пометкой вершин в дереве, а не полным удалением подтерма. С этого момента мы будем использовать обозначение $A_x[\lambda x. e]$ для терма A , у которого подтерм B , вершина “λx” и соответствующая вершина оператора применения помечены — вычеркнуты.

стым редексом. Более того, правила гарантируют, что выделенной может быть лишь вершина, находящаяся на самом левом пути терма. Иными словами, если система

переходов достигает конечного состояния, вычисления завершаются, нос-переменная выделена и не связана никаким простым редексом.

Пример. Рассмотрим терм $(\lambda x . x) @ (\lambda y . y)$.

$$\begin{aligned} \langle (\lambda x . x) @ (\lambda y . y); \quad []; \quad [] \rangle &\rightarrow^{[App]} \quad (2) \\ \langle (\lambda x . x) @ (\lambda y . y); \quad []; \quad [((\lambda y . y), [])] \rangle &\rightarrow^{[Lam-Elim]} \quad (3) \\ \langle \cancel{(\lambda x . x)} @ \cancel{(\lambda y . y)}; \quad [x \mapsto ((\lambda y . y), [])]; \quad [] \rangle &\rightarrow^{[BVar]} \quad (4) \\ \langle \cancel{(\lambda x . \lambda y . y)} @ \cancel{(\lambda y . y)}; \quad []; \quad [] \rangle &\rightarrow^{[Lam-Non-Elim]} \quad (5) \\ \langle \cancel{(\lambda x . \lambda y . y)} @ \cancel{(\lambda y . y)}; \quad []; \quad [] \rangle &\not\rightarrow \quad (6) \end{aligned}$$

Как и ожидалось, первым элементом конечного состояния является псевдоголовная нормальная форма входного терма: $(\lambda x . \lambda y . y) @ (\lambda y . y)$. Головной формой терма является терм $\lambda y . y$, в данном случае получаемый простым «выбрасыванием» вычеркнутых вершин из первого элемента конечного состояния системы переходов. В общем случае нам потребуется определить специальную функцию *exp*, которая строит соответствующий терм из конечного состояния, и показать её согласованность с головной редукцией (см. гл. 3.2). ■

Более показательный пример приведён на рис. 9.

$$exp \langle M[A]; \Gamma \bullet [(x \mapsto (B, \Gamma'))]; \Delta \rangle = exp \langle M[A[x/B[\Gamma']]]; \Gamma; \Delta \rangle \quad (7)$$

$$exp \langle M_B[A]; []; (B, \Gamma') : \Delta \rangle = exp \langle M_{B[\Gamma']}[A]; []; \Delta \rangle \quad (8)$$

$$exp \langle M; []; [] \rangle = M' \quad (9)$$

где $B[\Gamma'] = exp \langle B; \Gamma'; [] \rangle$, а терм M' получается из терма M удалением всех вычеркнутых вершин. (7) производит подстановку терма $B[\Gamma']$ вместо всех вхождений переменной x в такое поддерево терма M , которое имеет корнем выделенную вершину (т. е. поддерево A). Каждый рекурсивный вызов (8) производит подстановку всех переменных в соответствии с контекстом Γ' исключительно висячем аргументе B . Заметим, мы придерживаемся соглашения Барендрегта (Barendregt's convention): предполагается, что все переменные имеют уникальные имена. Следование этой концепции в данном случае позволяет избежать захвата имён переменных при подстановке. Далее мы будем называть *расширением* состояния системы переходов результат применения функции расширения *exp* к этому состоянию.

3.2 Согласованность с головной редукцией

В данной главе приводится доказательство согласованности головной линейной и головной редукций. Сначала определяется дополнительная функция *расширения* – *exp*.

Функция расширения. По данному состоянию системы переходов функция *exp* возвращает λ -терм. Интуитивно, функция *exp* производит последовательную редукцию всех простых редексов, накопившихся на данный момент времени. Поскольку последовательность простых редексов совпадает с последовательностью головных редексов, головная и линейная головная редукции согласованы. Формально¹³

Отметим некоторые важные для нас свойства системы переходов головной линейной редукции:

- Исходя из определения функции *exp* (конкретно уравнений (7), (8)), единственное правило системы переходов, при применении которого расширение получаемого состояния не равно расширению состояния, из которого оно получено, – [Lam-Elim], все остальные правила оставляют расширение неизменным.

- Количество переходов системы без применения правила [Lam-Elim] конечно. Данный факт является прямым следствием определения контекста, конечности размера входного терма и того, что в силу опре-

¹³ $\alpha \cdot \beta$ означает конкатенацию контейнеров α и β .

деления только правило [Lam-Elim] может изменить контекст. С этого момента мы будем обозначать последовательное применение правил без применения правила [Lam-Elim] через \rightarrow .

$$\begin{aligned} \text{exp}(2) &= \text{exp}((\lambda x . x) @ (\lambda y . y); []; []) = (\lambda x . x) @ (\lambda y . y) \\ \text{exp}(3) &= \text{exp}((\lambda x . x) @ (\lambda y . y); []; [(\lambda y . y), []]) = \text{exp}((\lambda x . x) @ (\lambda y . y); []; []) \\ &= (\lambda x . x) @ (\lambda y . y) \\ \text{exp}(4) &= \text{exp}(\cancel{(\lambda x . x)} @ \cancel{(\lambda y . y)}; [(x \mapsto ((\lambda y . y), []))]; []) \\ &= \text{exp}(\cancel{(\lambda x . \lambda y . y)} @ \cancel{(\lambda y . y)}; []; []) = \lambda y . y \\ \text{exp}(5) &= \text{exp}(\cancel{(\lambda x . \lambda y . y)} @ \cancel{(\lambda y . y)}; []; []) = \lambda y . y \\ \text{exp}(6) &= \text{exp}(\cancel{(\lambda x . \lambda y . y)} @ \cancel{(\lambda y . y)}; []; []) = \lambda y . y \end{aligned}$$

Идея доказательства согласованности головной и линейной головной редукций состоит в следующем: расширение не может быть изменено никаким правилом, кроме [Lam-Elim]. Следовательно, после применения функции расширения к состояниям системы переходов каждое применение правила [Lam-Elim] соответствует одному шагу головной редукции. Так, например, $(\lambda x . x) @ (\lambda y . y)$ соответствует первому шагу, $\lambda y . y$ соответствует трём оставшимся шагам.

Теорема 2. Пусть $\langle \dots \rangle$ – некоторое состояние системы переходов, расширение которого обозначено ... (см. иллюстрацию к теореме, рис. 7), такое, что на следующем шаге должно быть применено правило [Lam-Elim], и результатом его применения является состояние $\langle M_i; \Gamma_i; \Delta_i \rangle$, расширением которого является некоторый терм M'_i . Далее система переходов делает некоторое конечное число шагов до следующего применения правила [Lam-Elim], состояние $\langle M_i; \Gamma_i; \Delta_i \rangle$, расширением которого также является терм M'_i , тогда если может быть применено правило [Lam-Elim], расширяющее окружение Γ новым связыванием $(x \mapsto (B, \Gamma'))$, и расширением результата

Пример. В данном примере показан результат применения функции exp к каждому состоянию системы переходов для примера, приведённого в разделе 3.1.

применения которого является некоторый терм M'_{i+1} , то терм M'_{i+1} получается из термина M_i за одним шагом головной редукции.

Доказательство. Индукция по количеству применений правила [Lam-Elim].

База индукции. Очевидно, т. к. первый элемент, добавляемый в Γ , является головным редексом по определению.

Индукционный переход. Согласно индукционному предположению, после i -го применения правила [Lam-Elim] текущим является состояние вида $\langle M_i; \Gamma_i; \Delta_i \rangle$, расширением которого является некоторый терм M_i . Как уже отмечалось ранее, \rightarrow не изменяет расширения и число шагов \rightarrow конечно. Таким образом, существует два возможных случая:

1. Система переходов достигла конечного состояния, и нечего доказывать.

2. Система не достигла конечного состояния, и может быть применено правило [Lam-Elim]. В данном случае, согласно определению правила [Lam-Elim], мы знаем вид исходного состояния: $\langle A[\lambda x . e]; \Gamma; (B, \Gamma') : \Delta \rangle$, обозначим его за (i) , и вид результата применения правила [Lam-Elim] к нему: $\langle A_x[\cancel{\lambda x . e}]; (x \mapsto (B, \Gamma')) : \Gamma; \Delta \rangle$, обозначим

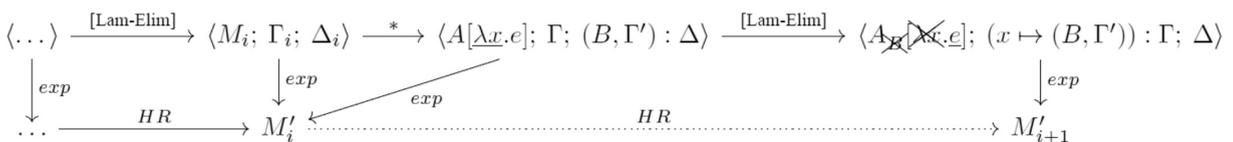


Рис. 7. Иллюстрация к теореме 2



его за (ii). Заметим, что нам надо показать, что результат применения функции exp к состоянию (ii) совпадает с результатом

$$exp \langle A[\underline{\lambda x}. e]; \Gamma; (B, \Gamma') : \Delta \rangle \quad (10)$$

$$\xrightarrow{* exp} exp \langle A[\underline{\lambda x}. e[\Gamma]]; []; (B, \Gamma') : \Delta \rangle \quad \text{за конечное число шагов, по (7)} \quad (11)$$

$$\xrightarrow{exp} exp \langle A_{B[\Gamma']}[\underline{\lambda x}. e[\Gamma]]; []; \Delta \rangle \quad \text{за один шаг (8)} \quad (12)$$

$$\xrightarrow{exp} \dots \quad (13)$$

Теперь применим функцию exp к состоянию (ii) :

$$exp \langle A_{\cancel{x}}[\cancel{\lambda x}. \underline{e}]; (x \mapsto (B, \Gamma')) : \Gamma; \Delta \rangle \quad (14)$$

$$\xrightarrow{* exp} exp \langle A_{\cancel{x}}[\cancel{\lambda x}. \underline{e}[\Gamma]]; [(x \mapsto (B, \Gamma'))]; \Delta \rangle \quad \text{устраняя } \Gamma, \text{ согласно (7)} \quad (15)$$

$$\xrightarrow{exp} exp \langle A_{\cancel{x}}[\cancel{\lambda x}. \underline{e}[\Gamma][x/B[\Gamma']]]; []; \Delta \rangle \quad \text{за шаг (7)} \quad (16)$$

$$\xrightarrow{exp} \dots \quad (17)$$

Легко заметить, что терм, являющийся первым компонентом состояния (12), $A_{B[\Gamma']}[\underline{\lambda x}. e[\Gamma]]$, имеет головным редексом $(\lambda x, B)$, поскольку контекст уже является пустым. Таким образом, если применить шаг головной редукции к данному терму, результатом будет терм $A_{\cancel{x}}[\cancel{\lambda x}. \underline{e}[\Gamma][x/B[\Gamma']]]$, по определению головной редукции, соответствующий первой компоненте состояния (16). Согласно (7) и (8), если продолжить применение функции exp к состояниям (12) и (16), соответственно, то в обоих случаях один и тот же стек висячих аргументов Δ не произведёт никаких изменений ни в аргументе B , ни в выделенном поддереве e . Следовательно, редекс $(\lambda x, B)$ является головным и для терма M_i , а значит, и расширением состояния (ii) является терм M_{i+1} , что и требовалось доказать ■

Теорема 2 сопоставляет каждому шагу головной редукции шаг системы переходов, соответствующий применению правила [Lam-Elim]. Таким образом, прямым следствием теоремы является тот факт, что *головная линейная редукция завершается тогда и только тогда, когда завершается головная редукция*. Ввиду того, что функция расширения не способна изменить путь от корня дерева терма до корня его выделенного поддеревя, первая компонента конечного состояния системы переходов для головной линейной редукции содержит

одного шага головной редукции терма M_i , т. е. термом M_{i+1} . Применим функцию exp к состоянию (i) :

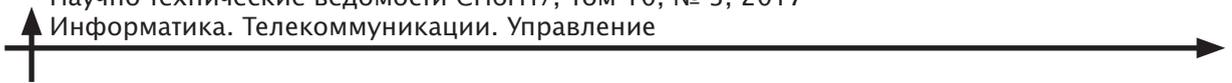
терм такой, что самый левый путь в дереве, его представляющем, является частью головной нормальной формы, а значит, и расширение конечного состояния является вершиной дерева Бёма. Таким образом, повторное применение головной линейной редукции ко всем оставшимся аргументам (полная головная линейная редукция, см. гл. 4) нормализует терм.

4 Полная головная линейная редукция

Полная головная линейная редукция (complete head linear reduction – CHLR) является расширением головной линейной редукции, рекурсивно применяющим последнюю к аргументам при достижении конечного состояния. В данной главе приводится формальное описание CHLR в виде системы переходов, являющейся истинным расширением системы переходов для HLR.

4.1 Система переходов для полной головной линейной редукции

Система переходов для CHLR отличается от системы переходов для HLR тремя новыми правилами [FVar-*]. Неформально говоря, эти правила обрабатывают ситуацию, когда система переходов HLR достигла своего конечного состояния, т. е. первым компонентом состояния системы переходов является некоторый терм с



Обозначения

$A[B]$	λ -терм с выделенным поддеревом B ¹¹
$\Gamma := (var \mapsto (t, \Gamma_1)) : \Gamma$	окружение (список)
$\Delta := (t, \Gamma) : \Delta$	стек висячих аргументов
$[]$	пустой список либо пустой стек
$\$$	символ-разделитель

Состояние $\langle A[B]; \Gamma; \Delta \rangle$	Начальное состояние $\langle A[A]; [], [] \rangle$	Конечное состояние $\langle A[x]; \Gamma; [] \rangle$
---	---	--

	Правила переходов	
$\langle A [e_1 @ e_2]; \Gamma; \Delta \rangle$	$\rightarrow \langle A [e_1 @ e_2]; \Gamma; (e_2, \Gamma) : \Delta \rangle$	[App]
$\langle A [\lambda x.e]; \Gamma; \$: \Delta \rangle$	$\rightarrow \langle A [\lambda x.e]; \Gamma; \$: \Delta \rangle$	[Lam-Non-Elim]
$\langle A [\lambda x.e]; \Gamma; (B, \Gamma') : \Delta \rangle$	$\rightarrow \langle A [\lambda x.e]; \Gamma; (x \mapsto (B, \Gamma')) : \Gamma; \Delta \rangle$	[Lam-Elim]
$\langle A [x]; [\dots, (x \mapsto (B, \Gamma')), \dots]; \Delta \rangle$	$\rightarrow \langle A [B]; \Gamma'; \Delta \rangle$	[BVar]
$\langle A [M [x] @ B]; \Gamma; (B, \Gamma') : \$: \Delta \rangle$	$\rightarrow \langle A [M [x] @ B]; \Gamma'; \$: \Delta \rangle,$ $x \notin dom(\Gamma)$	[FVar - 0]

NB: здесь B , как и во всех [FVar-*] правилах, является компонентом списка висячих аргументов. Более того, B должен быть аргументом некоторого вышестоящего в контексте A применения.

$\langle A [M [x] @ B]; \Gamma; (B, \Gamma') : C : \Delta \rangle$	$\rightarrow \langle A [M [x] @ B]; \Gamma'; \$: C : \Delta \rangle,$ $C \neq \$, x \notin dom(\Gamma)$	[FVar - 1]
$\langle A [M [x] @ B]; \Gamma; \$: (B, \Gamma') : \Delta \rangle$	$\rightarrow \langle A [M [x] @ B]; \Gamma'; \Delta_1 \rangle,$ $x \notin dom(\Gamma), \Delta_1 = \begin{cases} \Delta, & \text{если } \Delta = \$: \Delta_2 \\ \$: \Delta, & \text{иначе} \end{cases}$	[FVar - 2]

Рис. 8. Система переходов для полной головной линейной редукции

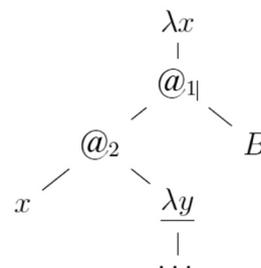
выделенной свободной или динамически несвязанной с аргументом переменной. Под *динамически несвязанной с аргументом* переменной понимается связанная переменная, чья абстракция не применена ни к какому аргументу. Итак, система переходов для CHLR обладает следующими изменениями по сравнению с системой переходов для HLR.

- Стек висячих аргументов Δ расширяется специальным символом $\$$. Символ $\$$ играет роль разделителя, запрещающего связывать абстракцию с аргументом, если путь от этого аргумента до соответствующей абстракции не является левым путём. Иными словами, разделитель $\$$ гарантирует, что все редексы в контексте являются простыми для данного подтерма. Например, пусть дан некоторый входной терм, и в ходе полной головной линейной редукции получено состояние $\langle M[\lambda y]; \Gamma; [\$, (B, \Gamma_0)] \rangle$, где терм

$M[\lambda y]$ изображён справа. В данном примере разделитель $\$$ запрещает связывание вершины λy с висячим аргументом B , поскольку они не образуют простого редекса.

- *Начальным* является состояние $\langle \lambda_1; []; [] \rangle$, где λ_1 является входным термом с выделенным корнем.

- *Конечным* состоянием является $\langle M[x]; \Gamma; [] \rangle$, где $x \notin dom(\Gamma)$. Заметим, что в отличие от головной линейной редукции, в конечном состоянии стек висячих аргументов должен быть пустым.



- Результатом CHLR является терм в нормальной форме; в терминах системы переходов это означает, что он может быть получен из первой компоненты конечного состояния системы переходов путём удаления из неё всех вычеркнутых вершин.

- На рис. 8 приведены правила для переходов для TS для {CHLR}. Для удобства чтения изменения по сравнению с соответ-

ствующими правилами для HLR выделены с помощью прямоугольников: выделение.

4.2 Функция расширения

Как и в случае HLR, мы определим функцию exp расширения состояния системы переходов до терма. В данном случае, по сравнению с системой переходов для HLR, добавляется лишь одно правило – (20).

$$exp \langle M[\underline{A}]; \Gamma \bullet (x \mapsto (B, \Gamma')); \Delta \rangle = exp \langle M[\underline{A}[x/B[\Gamma']]]; \Gamma; \Delta \rangle \quad (18)$$

$$exp \langle M_B[\underline{A}]; []; (B, \Gamma') : \Delta \rangle = exp \langle M_{B[\Gamma']}[\underline{A}]; []; \Delta \rangle \quad (19)$$

$$\boxed{exp \langle M_B[\underline{A}]; []; \$: \Delta \rangle} = \boxed{exp \langle M_B[\underline{A}]; []; \Delta \rangle} \quad (20)$$

$$exp \langle M; []; [] \rangle = M' \quad (21)$$

$$\text{где } B[\Gamma'] = exp \langle \underline{B}; \Gamma'; [] \rangle \quad (22)$$

$$M' \text{ получается из } M \text{ удалением вычеркнутых вершин} \quad (23)$$

Теорема 3. Полная головная линейная редукция завершается тогда и только тогда, когда терм имеет нормальную форму. Более того, если M – некоторый терм, $s_{final} ::= \langle M'; \Gamma'; \Delta' \rangle$ – конечное состояние системы переходов полной головной линейной редукции для терма M , обозначим за M_{exp} расширение состояния s_{final} , тогда:

1. M_{exp} есть M' за исключением вычеркнутых поддеревьев;

2. M_{exp} не содержит редексов;

3. M_{exp} является нормальной формой терма M .

Доказательство. Заметим, что правила [FVar-*] не могут изменить расширение состояния системы переходов. Следовательно, доказательство корректности CHLR является прямым следствием корректности HLR (см. следствия к теореме 2), а значит, полная головная линейная редукция завершается тогда и только тогда, когда завершается полная головная редукция терма,

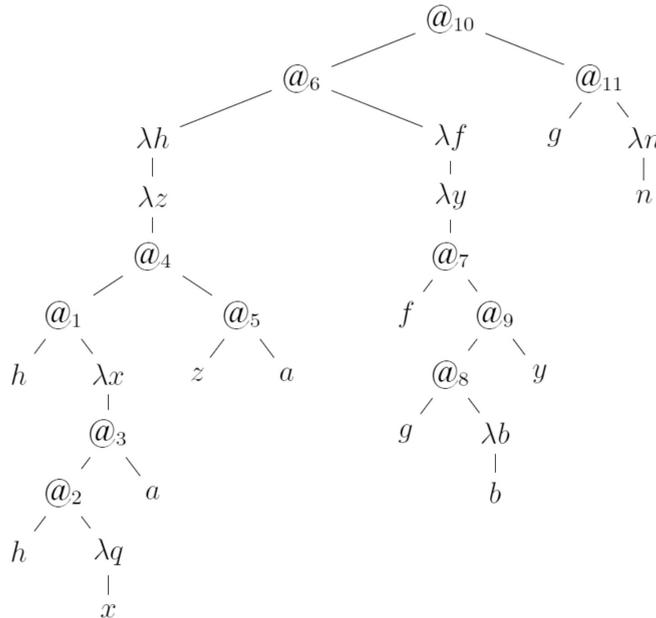


Рис. 9. Синтаксическое дерево терма $\langle NPR \rangle$

что в свою очередь эквивалентно конечности дерева Бёма, представляющего входной терм, а значит, и эквивалентно существованию нормальной формы входного терма. Иными словами, полная головная линейная редукция завершается тогда и только

тогда, когда у входного терма существует нормальная форма. ■

Далее мы рассмотрим пример применения системы переходов для головной линейной и полной головной линейной редукций соответственно на примере терма $\langle N P R \rangle$, где

$$N = \lambda h . \lambda z . h @ (\lambda x . (h @ (\lambda q . x) @ a)) @ (z @ a)$$

$$P = \lambda f . \lambda y . f @ ((g @ (\lambda b . b)) @ y)$$

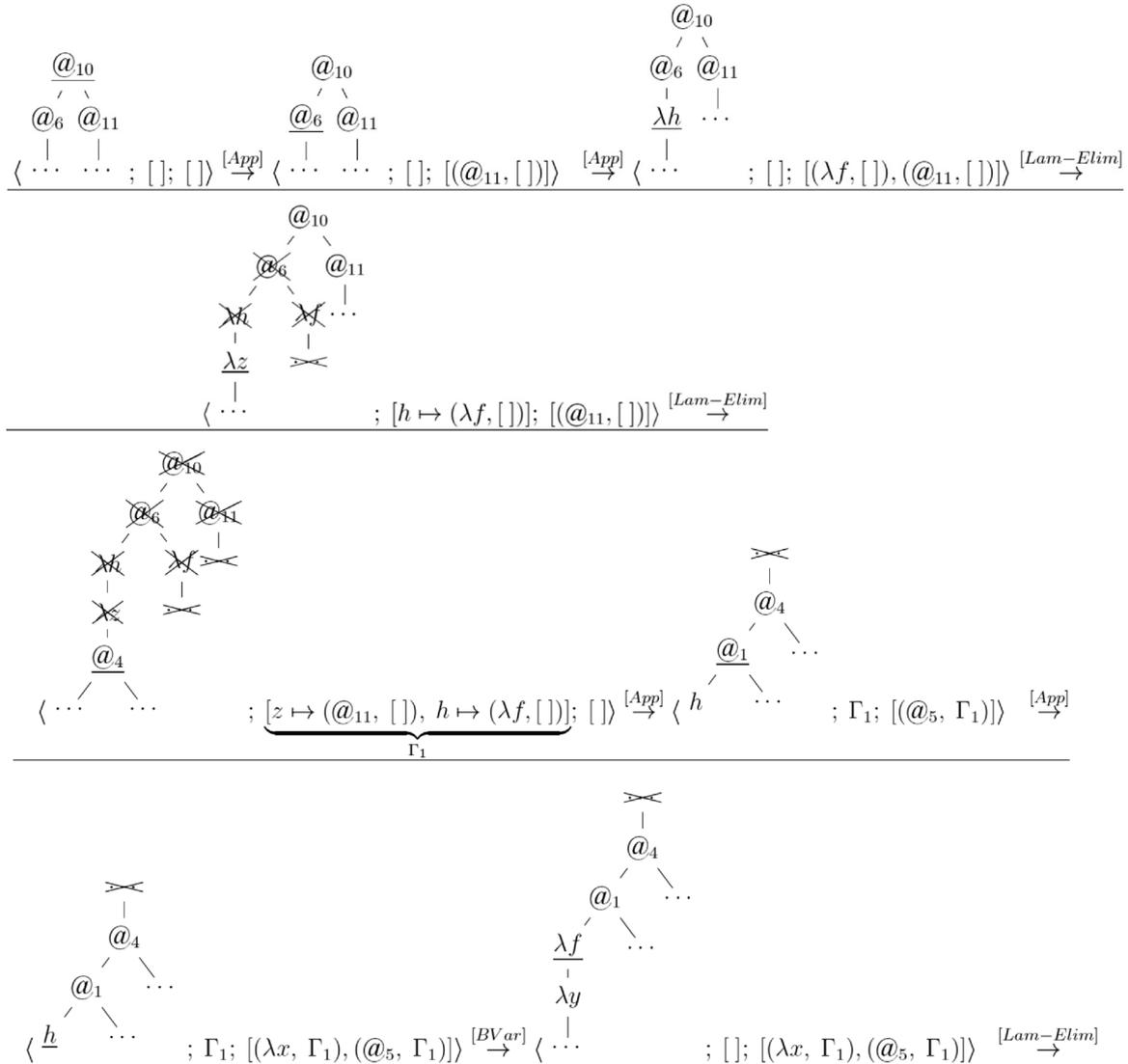
$$R = g @ (\lambda n . n)$$

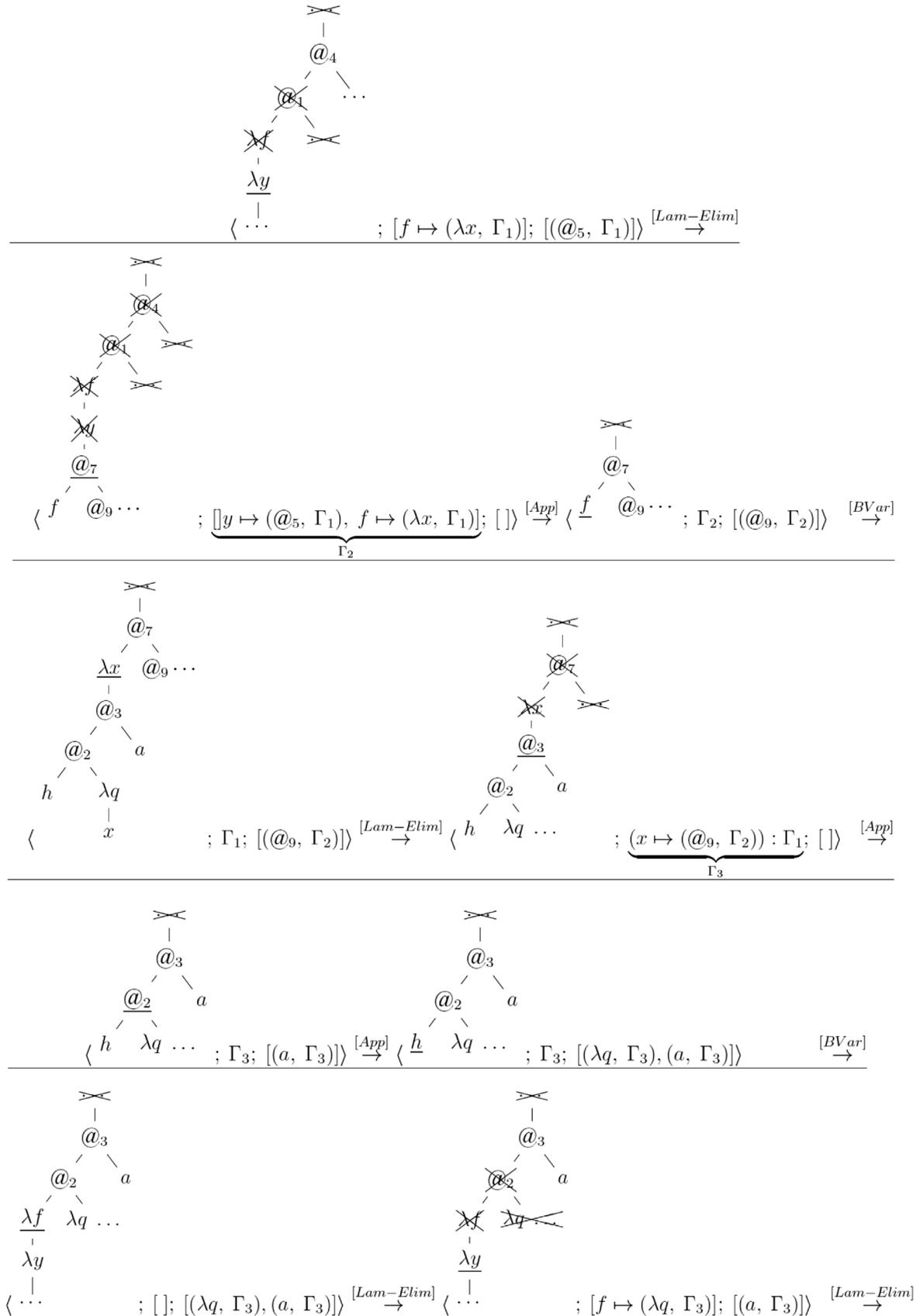
$$N P R = ((\lambda h \lambda z . (h @_1 (\lambda x . ((h @_2 (\lambda q . x)) @_3 a))) @_4 (z @_5 a)) @_6 (\lambda f \lambda y . f @_7 ((g @_8 (\lambda h . h)) @_9 y))) @_{10} (g @_{11} (\lambda n . n))$$

Для наглядности в нашем примере термы будут представляться частью их синтаксического

дерева. Синтаксическое дерево терма $\langle N P R \rangle$ может быть найдено на рис. 9.

Система переходов HLR для терма $\langle N P R \rangle$





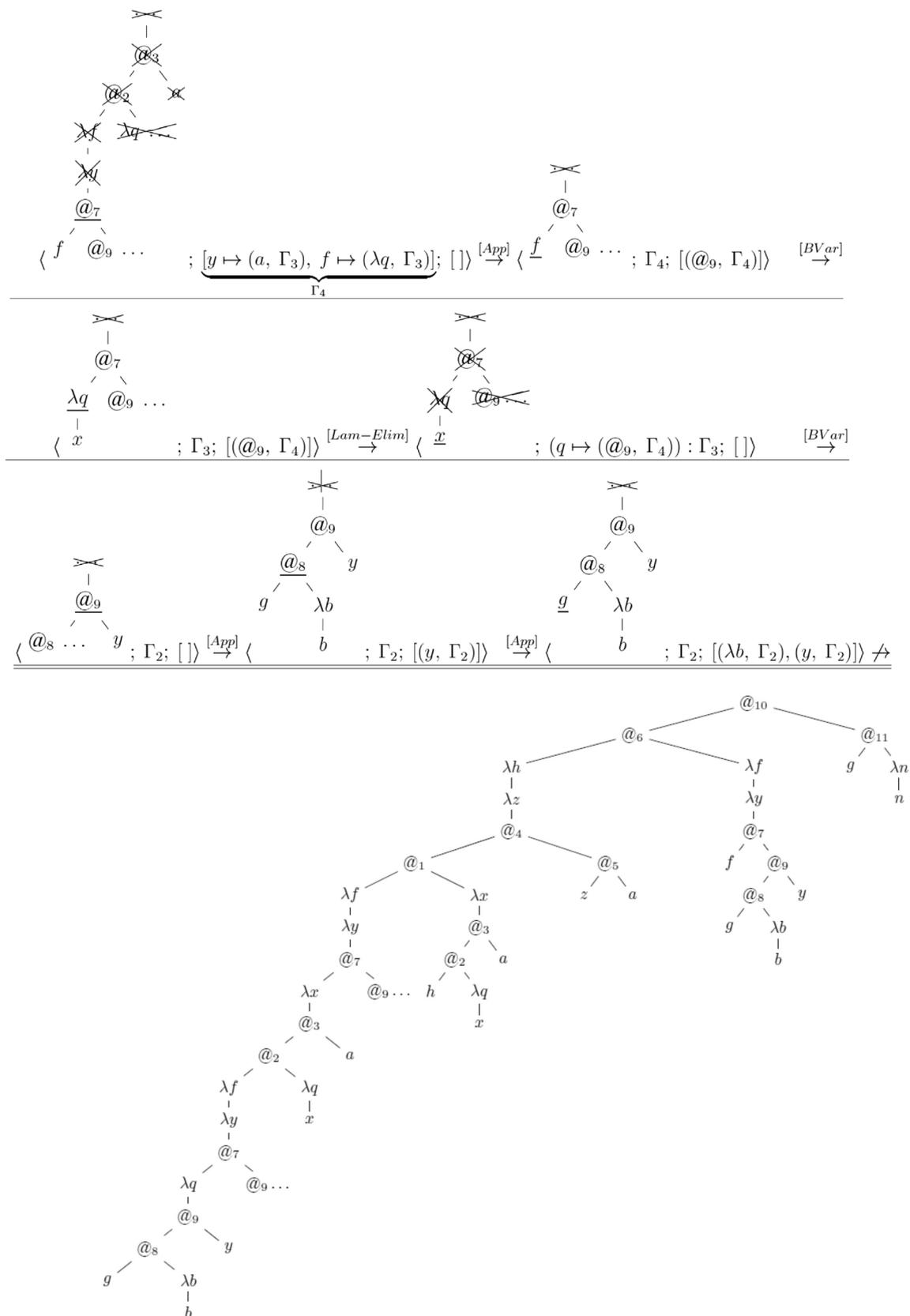
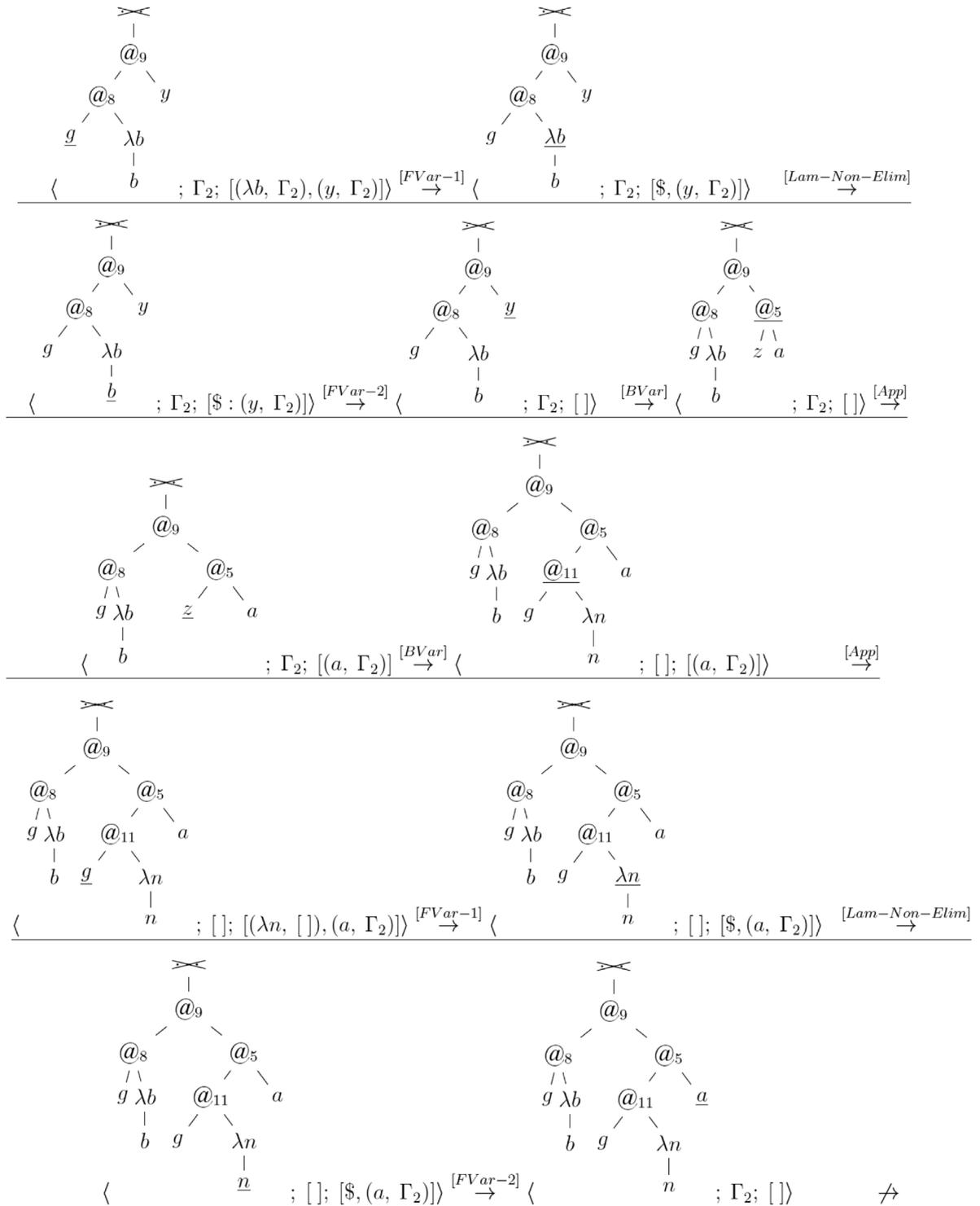


Рис. 10. Псевдоголовная нормальная форма (qhn) термина NPR

Замечание. Результатом головной линейной редукции является первый компонент конечного состояния системы переходов (NB: вычеркнутые элементы

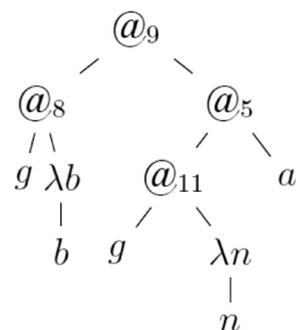
являются частью qhn). Таким образом, результатом головной линейной редукции является терм, представленный на рис. 10.

Система переходов для CHLR для терма $\langle NPR \rangle$



Заметим, что вплоть до первого применения одного из правил [FVar-*] система переходов для полной головной линейной редукции в точности повторяет шаги системы переходов для головной линейной редукции, поэтому мы просто продолжим построение нашего примера, начиная с конечного состояния системы переходов для головной линейной редукции.

Напомним, что результатом полной головной линейной редукции является первый компонент конечного состояния системы переходов, в котором все вычеркнутые вершины удалены:



СПИСОК ЛИТЕРАТУРЫ

1. Ong C.-H. Luke Normalisation by Traversals. CoRR, abs/1511.02629 // URL: <http://arxiv.org/abs/1511.02629>, 2015.
2. de Bruijn N.G. Lambda Calculus Notation with Nameless Dummies: A Tool for Automatic Formula Manipulation, with Application to the Church-Rosser Theorem // *Indagationes Mathematicae*. Elsevier, 1972. No. 34. Pp. 381–392.
3. Danos V., Regnier L. Head linear reduction. Unpublished, 2004.
4. Danos V., Herbelin H., Regnier L. Game semantics and abstract machines // Proc. of the 11th Annual IEEE Symp. on Logic in Computer Science. IEEE Computer Society, Washington, USA, 1996. 394 p.
5. Barendregt H.P. The lambda calculus: its syntax and semantics. Studies in logic and the foundations of mathematics. North-Holland, Amsterdam, New-York, Oxford, 1981.
6. Berezun D., Jones N.D. Compiling untyped lambda calculus to lower-level code by game semantics and partial evaluation (invited paper) // Proc. of the 2017 ACM SIGPLAN Workshop on Partial Evaluation and Program Manipulation. ACM, NY, USA, 2017. Pp. 1–11.
7. Sestoft P. Demonstrating lambda calculus re-

Заключение

Головная линейная редукция является основой для различных подходов к вычислениям [16, 20–25], в том числе и игровой семантики, из которой естественным образом вытекает процедура нормализации термов простого типизированного лямбда-исчисления [1]. Обобщение этого подхода до бестипового лямбда-исчисления позволяет определить новый подход к вычислениям – трассирующую нормализацию [1, 6].

В статье приведена формализация понятия головной линейной редукции для нетипизированного лямбда-исчисления в виде системы переходов, формально показана её корректность относительно головной редукции, впервые введено понятие полной головной линейной редукции, являющейся истинным обобщением головной линейной редукции, формализованное в виде системы переходов, а также показано, что последняя является эффективной редуцирующей стратегией, что является первым шагом в формальном доказательстве корректности трассирующей нормализации.

duction // *The Essence of Computation*. Springer-Verlag New York, Inc., 2002. Pp. 420–435.

8. Plotkin G.D. LCF considered as a programming language // *Theory of Computer Science*. 1977. No. 5(3). Pp. 223–255.
9. Keller R.M. Formal Verification of Parallel Programs // *Communications of the ACM*. 1976. Vol. 19. No. 7. Pp. 371–384.
10. Pierce B.C. Types and Programming Languages. The MIT Press, 2002.
11. Pierce B.C. Advanced Topics in Types and Programming Languages. The MIT Press, 2004.
12. Blum W. The safe lambda calculus. University of Oxford, UK, 2009.
13. Wadsworth Ch.P. Semantics and Pragmatics of the Lambda-Calculus. Oxford University, 1971.
14. François-Régis Sinot. Complete laziness: a natural semantics // *Electr. Notes Theor. Comput. Sci.* 2008. No. 204. Pp. 129–145.
15. Launchbury J. A natural semantics for lazy evaluation // In Conf. Record of the Twentieth Annual ACM SIGPLAN-SIGACT Symp. on Principles of Programming Languages. 1993. Pp. 144–154.
16. Jean-Jacques Levy. Optimal reductions in the lambda-calculus // *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formal-*



ism. Academic Press, 1980.

17. **Lamping J.** An algorithm for optimal lambda calculus reduction // Proc. of the 17th ACM SIGPLAN-SIGACT Symp. on Principles of Programming Languages. ACM, 1990. Pp. 16–30.

18. **Shivers O., Wand M.** Bottom-up beta-reduction: Uplinks and lambda-dags // Programming Languages and Systems. 14th European Symp. on Programming. 2005. Pp. 217–232.

19. **Kees-Lan can de Looij Vincent van Oostrom, Mariln Zwitterlood.** Lambdascope another optimal implementation of the lambda-calculus // Workshop on Algebra and Logic on Programming Systems. 2004.

20. **Mascari G., Pedicini M.** Head Linear Reduction and Pure Proof Net Extraction // Theor. Comput. Sci. 1994. Vol. 135. No. 1. Pp. 111–137.

21. **Baillet P., Pedicini M.** Elementary Complexity and Geometry of Interaction // Fundam. Inf. 2001. Vol. 45. No. 1-2. Pp. 1–31.

22. **Abramsky S., McCusker G.** Game semantics. In Computational Logic // Proc. of the 1997

Marktoberdorf Summer School. Springer Verlag, 1999. Pp. 1–56.

23. **Hyland J.M.E., Luke Ong C.-H.** On full abstraction for PCF: I, II, and III // Inf. Comput. 2000. Vol. 163(2). Pp. 285–408.

24. **Ker A.D., Nickau H., Ong C.-H. Luke** Innocent game models of untyped lambda-calculus // Theor. Comput. Sci. 2002. Vol. 272(1-2). Pp. 247–292.

25. **Blum W., Ong C.-H. Luke.** A concrete presentation of game semantics // Galop 2008: Games for Logic and Programming Languages. 2008.

26. **Abramsky S., Ong C.-H. Luke.** Full abstraction in the lazy lambda calculus // Inf. Comput. 1993. Vol. 105(2). Pp. 159–267.

27. **Ker A.D., Nickau H., Ong C.-H. Luke** A Universal Innocent Game Model for the Böhm Tree Lambda Theory // Computer Science Logic, 13th Internat. Workshop, 8th Annual Conf. of the EACSL. Madrid, Spain, 1999. Pp. 405–419.

28. **Gérard Huet.** Regular Böhm Trees // Math. Struct. in Comp. Science. 1998. No. 8. Pp. 671–680.

Статья поступила в редакцию 22.06.2017

REFERENCES

1. **Ong C.-H. Luke** *Normalisation by Traversals*. CoRR, abs/1511.02629. Available: <http://arxiv.org/abs/1511.02629>, 2015.

2. **de Bruijn N.G.** Lambda Calculus Notation with Nameless Dummies: A Tool for Automatic Formula Manipulation, with Application to the Church-Rosser Theorem. *Indagationes Mathematicae*. Elsevier, 1972, No. 34, Pp. 381–392.

3. **Danos V., Regnier L.** *Head linear reduction*. Unpublished, 2004.

4. **Danos V., Herbelin H., Regnier L.** Game semantics and abstract machines. *Proceedings of the 11th Annual IEEE Symposium on Logic in Computer Science*. IEEE Computer Society, Washington, DC, USA, 1996, 394 p.

5. **Barendregt H.P.** *The lambda calculus: its syntax and semantics*. *Studies in logic and the foundations of mathematics*. North-Holland, Amsterdam, New-York, Oxford, 1981.

6. **Berezun D., Jones N.D.** Compiling untyped lambda calculus to lower-level code by game semantics and partial evaluation (invited paper). *Proceedings of the 2017 ACM SIGPLAN Workshop on Partial Evaluation and Program Manipulation*. ACM, NY, USA, 2017, Pp. 1–11.

7. **Sestoft P.** *Demonstrating lambda calculus reduction*. *The Essence of Computation*. Springer-Verlag New York, Inc., 2002, Pp. 420–435.

8. **Plotkin G.D.** LCF considered as a programming language. *Theory of Computer Science*, 1977,

No. 5(3), Pp. 223–255.

9. **Keller R.M.** Formal Verification of Parallel Programs. *Communications of the ACM*, 1976, Vol. 19, No. 7, Pp. 371–384.

10. **Pierce B.C.** *Types and Programming Languages*. The MIT Press, 2002.

11. **Pierce B.C.** *Advanced Topics in Types and Programming Languages*. The MIT Press, 2004.

12. **Blum W.** *The safe lambda calculus*. University of Oxford, UK, 2009.

13. **Wadsworth Ch.P.** *Semantics and Pragmatics of the Lambda-Calculus*. Oxford University, 1971.

14. **François-Régis Sinot.** Complete laziness: a natural semantics. *Electr. Notes Theor. Comput. Sci.*, 2008, No. 204, Pp. 129–145.

15. **Launchbury J.** A natural semantics for lazy evaluation. *Conference Record of the Twentieth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, 1993, Pp. 144–154.

16. **Jean-Jacques Levy.** Optimal reductions in the lambda-calculus. *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*. Academic Press, 1980.

17. **Lamping J.** An algorithm for optimal lambda calculus reduction. *Proceedings of the 17th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*. ACM, 1990, Pp. 16–30.

18. **Shivers O., Wand M.** Bottom-up beta-reduction: Uplinks and lambda-dags. *Programming Languages and Systems, 14th European Symposium*

on Programming, 2005, Pp. 217–232.

19. **Kees-Lan van de Looij Vincent van Oostrom, Mariln Zwitterlood.** Lambdascope another optimal implementation of the lambda-calculus. *Workshop on Algebra and Logic on Programming Systems*, 2004.

20. **Mascari G., Pedicini M.** Head Linear Reduction and Pure Proof Net Extraction. *Theor. Comput. Sci.*, 1994, Vol. 135, No. 1, Pp. 111–137.

21. **Baillet P., Pedicini M.** Elementary Complexity and Geometry of Interaction. *Fundam. Inf.*, 2001, Vol. 45, No. 1-2, Pp. 1–31.

22. **Abramsky S., McCusker G.** Game semantics. In Computational Logic. *Proceedings of the 1997 Marktoberdorf Summer School*. Springer Verlag, 1999. Pp. 1–56.

23. **Hyland J.M.E., Luke Ong C.-H.** On full abstraction for PCF: I, II, and III. *Inf. Comput.*, 2000,

Vol. 163(2), Pp. 285–408.

24. **Ker A.D., Nickau H., Ong C.-H. Luke** Innocent game models of untyped lambda-calculus. *Theor. Comput. Sci.*, 2002, Vol. 272(1-2), Pp. 247–292.

25. **Blum W., Ong C.-H. Luke.** A concrete presentation of game semantics. *Galop 2008: Games for Logic and Programming Languages*, 2008.

26. **Abramsky S., Ong C.-H. Luke.** Full abstraction in the lazy lambda calculus. *Inf. Comput.*, 1993, Vol. 105(2), Pp. 159–267.

27. **Ker A.D., Nickau H., Ong C.-H. Luke** A Universal Innocent Game Model for the Böhm Tree Lambda Theory. *Computer Science Logic, 13th International Workshop, 8th Annual Conference of the EACSL*, Madrid, Spain, 1999, Pp. 405–419.

28. **Gérard Huet.** Regular Böhm Trees. *Math. Struct. in Comp. Science*, 1998, No. 8, Pp. 671–680.

Received 22.06.2017

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

Березун Даниил Андреевич

Berezun Daniil A.

E-mail: d.berezun@2009.spbu.ru



Моделирование вычислительных,
телекоммуникационных, управляющих
и социально-экономических систем

DOI: 10.18721/JCSTCS.10307

УДК 004.896

**ПРИНЦИПЫ НЕГОЛОНОМНОЙ МЕХАНИКИ В УПРАВЛЕНИИ
ДВИЖЕНИЕМ ГРУППЫ ТРАНСПОРТНЫХ РОБОТОВ**

А.С. Смирнов¹, Б.А. Смольников¹, В.А. Леонтьев²

¹Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация;

²Центральный научно-исследовательский и опытно-конструкторский институт
робототехники и технической кибернетики, Санкт-Петербург, Российская Федерация

Обсуждены вопросы централизованного управления движением отдельного минипоезда, входящего в состав группы роботов и состоящего из 5-6 грузовых тележек, шарнирно связанных друг с другом. Эти тележки образуют своеобразный «неголономный хвост», боковые перемещения которого с трудом поддаются контролю и корректировке, тем самым представляя опасность для окружающих членов группы. Кроме ухудшения ходовой компактности, эти боковые колебания хвоста существенно снижают устойчивость движения цепочки тележек и даже влияют на динамику движения и маневрирования головного модуля. Для оценки этих опасностей и выбора предельно допустимой скорости движения минипоезда в статье построена и проанализирована математическая модель движения неголономной цепочки, позволяющая центральному оператору прогнозировать форму и размеры опасной зоны отдельного минипоезда, тем самым предотвращая возникновение нештатных ситуаций.

Ключевые слова: группа роботов; минипоезд; неголономный хвост; головной модуль; централизованное управление.

Ссылка при цитировании: Смирнов А.С., Смольников Б.А., Леонтьев В.А. Принципы неголономной механики в управлении движением группы транспортных роботов // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2017. Т. 10. № 3. С. 83–91. DOI: 10.18721/JCSTCS.10307

**PRINCIPLES OF NONHOLONOMIC MECHANICS IN CONTROLLING
A TRANSPORT ROBOT GROUP**

A.S. Smirnov¹, B.A. Smolnikov¹, V.A. Leontev²

¹Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation;

²State Scientific Center for Robotics and Technical Cybernetics (RTC),
St. Petersburg, Russian Federation

The problems of centralized control of motion of a separate mini-train that is part of a robot group are discussed. Each mini-train consists of 5-6 trolleys jointly

connected one by one. These trolleys are a kind of a «nonholonomic tail» whose lateral movements are difficult to control and adjust, thus being a danger to other members of the group. These lateral oscillations worsen the trajectory compactness, and they also significantly reduce the stability of the whole chain and affect the dynamics and maneuvering of the heading module. To estimate these dangers and select the maximum permissible speed of the mini-train, a mathematical model of motion of such nonholonomic chain is built and analyzed in the article. It allows the central operator to predict the shape and dimensions of the danger zone for a separate mini-train, and to therefore prevent the occurrence of abnormal situations.

Keywords: robot group; mini-train; nonholonomic movement; stability of movement; centralized control.

Citation: Smirnov A.S., Smolnikov B.A., Leontev V.A. Principles of Nonholonomic Mechanics in Controlling a Transport Robot Group. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems. 2017, Vol. 10, No. 3, Pp. 83–91. DOI: 10.18721/JCSTCS.10307

Введение

Постепенное внедрение разнообразных роботов в реальную жизнь современного общества поднимает множество проблем их рационального использования в различных ситуациях, причем эти проблемы не только носят экономический характер, но и порождают вопросы выделения сфер их применения, разработки специальных правил общения и/или обращения с ними и т. д. [1, 2]. Особенно важную роль эти вопросы сыграют в тех ситуациях, когда вместо одиночных роботов различных конструкций и различного назначения будут создаваться для решения тех или иных задач специализированные группы роботов, содержащие либо несколько роботов, либо несколько десятков роботов [3]. Именно такие группы смогут стать серьезным стимулом экономического или военного прогресса, и именно здесь на первый план выйдут требования к формированию рациональных принципов эксплуатации групп роботов и законов управления ими [4, 5].

Следует подчеркнуть, чем отличается в смысле управляемости группа роботов от группы живых организмов. Основные отличия в «мышлении» роботов по сравнению с мышлением живых организмов состоят в том, что они лишены чувства страха перед уничтожением или внезапным повреждением (разумеется, в пределах заданной программы поведения), хотя в случае крайней необходимости могут совершать такие действия, как самоподрыв или иной экстремальный акт. Кроме того, в случае

возникновения любых непредвиденных обстоятельств робот не может столь же целесообразно найти рациональный выход из сложившейся ситуации, как это может сделать человек [6]. Особенно заметно это различие в поведении проявляется при групповом использовании людей или роботов [7]. В группе людей-исполнителей каждый индивидуум выбирает принцип локальной оптимизации своих действий, тогда как в группе роботов-исполнителей, обладающих менее совершенной и менее универсальной сенсорикой, их действия должны контролироваться и координироваться неким центральным оператором, обладающим глобальной информацией о текущем расположении и состоянии группы и окружающей среды [8]. Эта глобальность является важным преимуществом группы роботов перед группой людей, однако она требует и более четкого управления со стороны центрального оператора.

Разумеется, конкретный режим группового управления определяется заданными целями и ресурсами проводимой операции, однако в ней всегда первостепенную роль играет тесное и целенаправленное взаимодействие отдельных роботов [9]. Характер этого взаимодействия зависит как от конструктивных особенностей роботов, так и от типа совместно выполняемых операций [10]. Именно эти факторы определяют идеологию и технологию процессов управления групповым поведением. Следует также подчеркнуть тесную взаимосвязь алгоритмов управления с конструктивными осо-

бенностями роботов, образующих группу, т. к. именно эти особенности нередко определяют как стратегию, так и тактику действий центрального оператора [11, 12].

Постановка задачи

Обратимся к рассмотрению весьма специфической группы транспортных роботов, используемых в цехах крупных автомобильных заводов [13, 14]. Основная функция этих роботов – развозка деталей и устройств по автосборочным линиям и конвейерам [15]. В силу большого разнообразия и большого количества доставляемых деталей, транспортные роботы представляют собой подобие автопоездов с электротягой, снующих по огромной цеховой территории по самым разнообразным трассам и с разными скоростями [16]. Головной модуль каждого минипоезда снабжен системой управления и аппаратурой наблюдения за окружающей обстановкой. Прицепленные к нему грузовые тележки не имеют собственной системы управления и образуют пассивную неголономную цепочку, состоящую из 5-6 отдельных звеньев.

Рассматривая этот минипоезд как отдельного робота с весьма длинной хвостовой частью, отметим, что он представляет собой нестандартно управляемый неполноприводный объект, обладающий в процессе движения переменной конфигурацией, которая зависит от режима движения головного модуля, и испытывающий серьезные проблемы с устойчивостью движения по некоторым траекториям. Под переменной конфигурацией здесь понимается переменная геометрия ломаной линии, образованной межшарнирными отрезками цепочки тележек.

Эти проблемы приобретают особую важность, если учесть, что на существующих крупных автозаводах число подобных транспортных минипоездов, перемещающихся в одном цеху, может достигать 100 единиц и более. В таких условиях во избежание их столкновений необходимо гарантировать четкую динамическую конфигурацию хвостовой части каждого робота, поведение которой при неполноприводном управлении требует повышенного внима-

ния центрального оператора.

Цель настоящей статьи – построение и анализ математической модели неголономной цепочки грузовых тележек, следующих за тяговым модулем, а также прогнозирование текущей конфигурации цепочки с учетом неголономных связей.

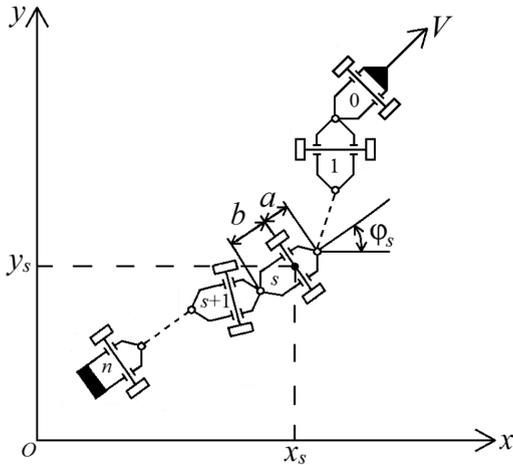
Построение и анализ математической модели

Чтобы выяснить возникающие здесь неблагоприятные ситуации и оценить диапазон поперечных колебаний хвостовой части робота, рассмотрим упрощенную кинематическую картину этих колебаний, обусловленную неголономным характером движений каждой грузовой тележки (т. е. каждого звена цепочки) [17, 18]. Обозначая индексом s номер s -го модуля этой цепочки, положим, что головной (т. е. тяговый) модуль имеет $s = 0$, так что декартовы координаты его центра (под которым будем понимать точку пересечения его продольной оси с осью качения его тяговой колесной пары) будут x_0, y_0 , а угол φ_0 будет углом между его продольной осью и осью Ox . Аналогичные обозначения примем и для прочих модулей цепочки (см. рисунок).

Положим теперь, что первый модуль по указанию центрального оператора совершает заданное движение по плоскости Oxy , а остальные модули, шарнирно связанные с ним и между собой, образуют протяженную кинематическую цепочку, звенья которой связаны между собой как голономными, так и неголономными связями [19]. Примем, что все звенья совершенно одинаковы, имеют длину между шарнирами $a+b$ и положение s -го модуля на плоскости xy определяется обобщенными координатами x_s, y_s, φ_s , где x_s, y_s – декартовы координаты его центра (точки пересечения продольной оси звена с осью колесной пары), а φ_s – угол между продольной осью звена и осью x . Ясно, что между обобщенными координатами соседних модулей существуют следующие соотношения связи:

$$\begin{aligned} x_{s+1} &= x_s - b \cos \varphi_s - a \cos \varphi_{s+1}; \\ y_{s+1} &= y_s - b \sin \varphi_s - a \sin \varphi_{s+1}. \end{aligned} \quad (1)$$

Кроме этих голономных связей суще-



Цепочка грузовых тележек

ствуют также и неголономные связи, выражающие требование коллинеарности вектора скорости центра s -го модуля с его продольной осью [20, 21]:

$$\dot{x}_s \sin \varphi_s - \dot{y}_s \cos \varphi_s = 0. \quad (2)$$

Далее будем считать, что головной модуль, отвечающий $s = 0$, является тягачом и совершает некоторое предписанное движение по плоскости xy , тогда как остальные n звеньев образуют ее «неголономный хвост», характер движения которого мы и хотим изучить.

Подставляя теперь в соотношение (2) выражения для x_s, y_s , согласно (1) приходим к следующей системе уравнений:

$$\begin{aligned} & \dot{x}_0 \sin \varphi_0 - \dot{y}_0 \cos \varphi_0 = 0 \\ & \dot{x}_0 \sin \varphi_1 - \dot{y}_0 \cos \varphi_1 + \\ & + b\dot{\varphi}_0 \cos(\varphi_0 - \varphi_1) + a\dot{\varphi}_1 = 0 \\ & \dots \dots \dots (3) \\ & \dot{x}_0 \sin \varphi_n - \dot{y}_0 \cos \varphi_n + \\ & + b\dot{\varphi}_0 \cos(\varphi_n - \varphi_0) + (a + b)\dot{\varphi}_1 \cos(\varphi_0 - \varphi_1) + \\ & + (a + b)\dot{\varphi}_2 \cos(\varphi_n - \varphi_2) + \dots \\ & \dots + (a + b)\dot{\varphi}_{n-1} \cos(\varphi_1 - \varphi_{n-1}) + a\dot{\varphi}_n = 0. \end{aligned}$$

Если считать, что в этой системе уравнений координаты тягового модуля заданы как функции времени $x_0(t), y_0(t)$, то угол φ_0 найдется из первого уравнения, тогда как остальные n уравнений будут вполне достаточны для нахождения n переменных $\varphi_1, \varphi_2, \dots, \varphi_n$. Поэтому систему (3) можно

с полным правом назвать системой уравнений движения рассматриваемой неголономной цепочки. При этом, как видно, отпадает необходимость использования динамических свойств цепочки, а, следовательно, и динамических уравнений.

Таким образом, весь процесс движения цепочки предопределен движением ее головного звена и условиями неголономных связей. Интересно отметить, что согласно стандартному определению числа степеней свободы механической системы (число обобщенных координат минус число уравнений связей) у данной цепочки оно оказывается равным нулю, что противоречит здравому смыслу [22]. Уместно поэтому считать, что рассматриваемая цепочка имеет только кинематические степени свободы, не подпадающие под приведенное определение, которое следует относить к механическим системам с явно выраженными динамическими свойствами.

Построенная система (3) представляет собой нелинейную систему дифференциальных уравнений, имеющую цепную структуру и допускающую возможность последовательного интегрирования уравнений первого порядка. Второй особенностью этой системы является тот факт, что движение $(s+1)$ -го модуля цепочки полностью предопределено движением ее s -го модуля (это следует из исходных соотношений (1), (2)). Поэтому если какому-либо заданному режиму движения головного модуля отвечает аналогичный режим движения следующего за ним первого модуля, то и движения всех последующих модулей будут также происходить в подобном режиме, хотя возможно, с несколько измененными параметрами.

Чтобы выявить основные свойства изучаемой системы, рассмотрим некоторые простейшие режимы ее движения.

Пусть, например, головной модуль совершает равномерное движение по окружности радиуса R . Нетрудно понять, что после некоторого переходного процесса и все прочие модули цепочки выйдут на круговые траектории, образовав равномерно вращающуюся спираль. Геометрия этой спирали определится, очевидно, последовательно-

стью межзвенных углов $\psi_s = \varphi_{s-1} - \varphi_s$, а также последовательностью радиусов траекторных окружностей $\rho_s = \sqrt{x_s^2 + y_s^2}$. Для нахождения этих последовательностей положим, что

$$\begin{aligned} x_0(t) &= R \sin \Omega t & y_0(t) &= -R \cos \Omega t \\ \dot{x}_0(t) &= \Omega R \cos \Omega t & \dot{y}_0(t) &= \Omega R \sin \Omega t. \end{aligned} \quad (4)$$

Подставляя эти значения в (3), найдем $\varphi_1 = \Omega t$. Для угла ψ_1 получим уравнение

$$b \cos \psi_1 - R \sin \psi_1 = -a, \quad (5)$$

решение которого есть

$$\operatorname{tg} \frac{\psi_1}{2} = \frac{a + b}{R + \sqrt{R^2 + b^2 - a^2}}. \quad (6)$$

Отсюда следует необходимость выполнения ограничения на параметры спирали:

$$a \leq \sqrt{R^2 + b^2}. \quad (7)$$

Соответствующий радиус окружности ρ_1 согласно (1) представится выражением:

$$\begin{aligned} \rho_1^2 &= R^2 + a^2 + b^2 + 2ab \cos \psi_1 - 2bR \sin \psi_1 = \\ &= R^2 + b^2 - a^2, \end{aligned} \quad (8)$$

из которого становится понятным смысл ограничения (7). Видно также, что если «хвостовая» часть модуля превышает его «головную» часть, т. е. $b > a$, то $\rho_1 > R$ и, следовательно, $\rho_{s+1} > \rho_s$, так что спираль будет раскручиваться с ростом номера s . Если же $b < a$, то спираль будет скручиваться. Наконец, при $b = a$ все модули будут двигаться по начальной окружности радиуса R .

Чтобы выявить геометрические свойства полученной спирали, положим $b/a = \beta$ и запишем соотношения (6) и (8) в виде

$$\rho_s^2 - \rho_{s-1}^2 = a^2(\beta^2 - 1), \quad \operatorname{tg} \frac{\psi_1}{2} = \frac{\rho_s - \rho_{s-1}}{a(\beta - 1)}. \quad (9)$$

Видно, что при $a \ll \rho_s$ второе из написанных выражений может быть представлено в виде

$$\Delta \rho = \frac{1}{2} a(\beta - 1) \Delta \varphi, \quad (10)$$

показывающем, что приращение радиуса спирали оказывается пропорциональным приращению полярного угла. Этим свойством, как известно, обладает спираль

Архимеда, к которой и будет стремиться рассматриваемая цепочка по мере роста радиуса ρ_s .

Исследуем теперь устойчивость прямолинейного движения цепочки. С этой целью, примем в системе (3), что $\dot{x}_0 = V, \dot{y}_0 = 0$, и линеаризуя ее вблизи невозмущенного движения $\varphi_s = 0$, представим ее в виде ($v = V/a$):

$$\begin{aligned} \dot{\varphi}_1 + v\varphi_1 &= 0 \\ (1 + \beta)\dot{\varphi}_1 + \dot{\varphi}_2 + v\varphi_2 &= 0 \\ (1 + \beta)\dot{\varphi}_1 + (1 + \beta)\dot{\varphi}_2 + \dot{\varphi}_3 + v\varphi_3 &= 0 \quad (11) \\ \dots\dots\dots \\ (1 + \beta)\dot{\varphi}_1 + (1 + \beta)\dot{\varphi}_2 + \dots + \dot{\varphi}_n + v\varphi_n &= 0. \end{aligned}$$

Видно, что характеристический определитель этой системы имеет треугольную структуру, так что отвечающее ему характеристическое уравнение

$$\Delta(p) = (p + v)^n = 0 \quad (12)$$

имеет n -кратный, вещественный отрицательный корень, гарантирующий асимптотическую устойчивость прямолинейной траектории. Тем не менее, высокая кратность этого корня (при достаточно большом n) говорит о том, что частные решения системы (11) будут иметь следующий вид:

$$\begin{aligned} \varphi_s &= e^{-vt} [c_1 + c_2(vt) + c_3(vt)^2 + \dots \\ &\dots + c_s(vt)^{s-1}], \end{aligned} \quad (13)$$

где c_1, \dots, c_s — константы. Ясно, что эти решения будут при $t \rightarrow \infty$ обращаться в ноль, однако в силу своей немонотонности могут проходить через весьма значительные максимальные значения. Это означает, что с ростом s уровень возмущений нарастает, ввиду чего хвостовые модули испытывают более значительные отклонения от невозмущенной траектории, чем модули, близкие к тягачу. Этот эффект хорошо известен на практике и, как видно, в его возникновении существенную роль играют неголономные свойства цепочки.

В заключение рассмотрим наиболее интересный синусоидальный режим движения головного модуля, когда

$$\begin{aligned} \dot{x}_0 &= V = \text{const}, \quad y_0 = a\delta \sin \Omega t, \\ \dot{y}_0(t) &= a\Omega\delta \cos \Omega t, \end{aligned} \quad (14)$$

где δ – безразмерная амплитуда поперечных смещений головного модуля.

Полагая, что при $\delta \ll 1$ углы φ_s также остаются достаточно малыми, линеаризуем систему (3), приведя ее к виду:

$$\begin{aligned} V\varphi_0 &= a\Omega\delta \cos \Omega t, \\ a\dot{\varphi}_1 + V\varphi_1 &= a\Omega\delta \cos \Omega t - b\dot{\varphi}_0 \\ a\dot{\varphi}_2 + V\varphi_2 &= a\Omega\delta \cos \Omega t - b\dot{\varphi}_0 - (a+b)\dot{\varphi}_1 \end{aligned} \quad (15)$$

Эта система решается путем последовательной интеграции цепочки линейных уравнений, в результате чего находим:

$$\begin{aligned} \varphi_0 &= \frac{\Omega\delta}{v} \cos \Omega t \\ \varphi_1 &= \frac{\Omega\delta}{v(\Omega^2 + v^2)} [\Omega v(1 + \beta) \sin \Omega t + \\ &+ (v^2 - \beta\Omega^2) \cos \Omega t]. \end{aligned} \quad (16)$$

Записывая решение уравнения для угла φ_1 в виде

$$\begin{aligned} \varphi_1 &= \frac{\Omega\delta}{v(\Omega^2 + v^2)} \times \\ &\times \sqrt{(v^2 - \beta\Omega^2)^2 + \Omega^2 v^2 (1 + \beta)^2} \cos(\Omega t - \gamma_1), \quad (17) \\ \operatorname{tg} \gamma_1 &= \frac{\Omega v(1 + \beta)}{v^2 - \beta\Omega^2}, \end{aligned}$$

замечаем, что оно по своему характеру повторяет колебательное движение головного модуля, хотя и с измененной амплитудой и с некоторым запаздыванием по фазе. Этот сдвиг фазы характеризуется углом γ_1 , который может принимать различные значения в зависимости от соотношения параметров v, β, Ω . В частности, при $\beta = v^2/\Omega^2$ этот сдвиг оказывается равным $\pi/2$.

Более важным является отношение амплитуды угловых колебаний первого модуля к амплитуде колебаний головного. Для оценки этого отношения введем коэффициент

$$K_{01} = \left(\frac{\operatorname{am} \varphi_1}{\operatorname{am} \varphi_0} \right)^2 = \frac{v^2 + \beta^2 \Omega^2}{\Omega^2 + v^2}, \quad (18)$$

величина которого, очевидно, существенно зависит от β . Так, если $\beta = 1$, то $K_{01} = 1$, и амплитуда колебаний модуля оказывается равной амплитуде тягача. При $\beta > 1$ или $\beta < 1$ происходит ее увеличение или уменьшение соответственно. Нетрудно показать,

что аналогичный характер носят и поступательные движения этого модуля. Так, линеаризуя второе из соотношений (1), найдем из него:

$$y_1 = y_0 - b\varphi_0 - a\varphi_1 = \frac{a\delta}{\Omega^2 + v^2} \times \sqrt{(v^2 - \beta\Omega^2)^2 + \Omega^2 v^2 (1 + \beta)^2} \sin(\Omega t - \gamma_1). \quad (19)$$

Легко видеть, что и для y_1 имеют место те же соотношения связи с y_0 , как и для φ_0 с φ_1 . Отсюда следует, что нет необходимости строить решения для последующих модулей, т. к. их движения будут качественно повторять движения головного модуля.

Таким образом, можно сделать вывод о том, что синусоидальное движение головного модуля приводит к аналогичному движению всех последующих модулей неголономной цепочки, причем амплитуды этих синусоид будут либо нарастать с ростом s (при $\beta > 1$), либо убывать (при $\beta < 1$), либо оставаться неизменными по всей длине цепочки (при $\beta = 1$).

Заключение

Рассмотренный пример показывает, что неголономные механические системы могут обнаруживать весьма нетривиальное поведение, совершенно не зависящее от их динамических и статических характеристик, но явно связанное с их геометрическими параметрами. В представленной неголономной цепочке таким ключевым параметром служит степень геометрической асимметрии модуля – параметр $\beta = b/a$, который и определяет степень опасности боковых колебаний хвостового модуля в группе транспортных роботов. Эта опасность связана не только с кинематическими эффектами неголономного движения транспортного робота, но и с сопровождающими их динамическими перегрузками, которые могут вызывать, например, опрокидывание хвостовых модулей, что нередко наблюдается на практике.

Построенная математическая модель позволяет центральному оператору предвидеть и предупреждать возникновение подобных ситуаций. Для этого оператору следует группировать движущиеся мини-

поезда в некоторые «струи», чтобы избежать их близости со встречными потоками, а также снизить кривизну их траекторий. Естественно, что важнейшими параметрами этих «струй» окажутся средняя скорость движения минипоездов, допустимая плотность их распределения в струе, длина тормозного пути в случае необходимости

остановки, а также показатель их путевой маневренности (например, минимальный радиус разворота). Совместный учет перечисленных факторов приводит к проблеме поиска их оптимальных значений, обеспечивающих минимизацию числа минипоездов при заданных требованиях к объему перевозок.

СПИСОК ЛИТЕРАТУРЫ

1. Юревич Е.И., Новаченко С.И., Павлов В.А. Управление роботами от ЭВМ. М.: Наука, 1980. 261 с.
2. Макаров И.М., Рахманкулов В.З. Групповое управление роботами-манипуляторами с распределенно-централизованной организацией обработки информации // Микропроцессорные системы управления в радиотехнике. М.: Наука, 1984. С. 35–45.
3. Каляев И.А., Гайдук А.Р., Капустян С.Г. Модели и алгоритмы коллективного управления в группах роботов. М.: Физматлит, 2009. 280 с.
4. Строев В. Системы с искусственным интеллектом в сухопутных войсках // Зарубежное военное обозрение. 1997. № 3. С. 27–30.
5. Рубцов И.В., Нестеров В.Е., Рубцов В.И. Современная зарубежная военная микро- и мини-робототехника // Микросистемная техника. 2000. № 3. С. 36–42.
6. Каляев И.А., Гайдук А.Р., Капустян С.Г. Распределенные системы планирования действий коллективов роботов. М.: Янус-К, 2002. 291 с.
7. Ширяев В.И. Об управлении коллективом роботов при игре в футболе как задаче управления в условиях неточной информации // Искусственный интеллект. 2000. № 3. С. 353–360.
8. Щербатов И.А., Проталинский И.О., Проталинский О.М. Управление группой роботов: компонентный подход // Информатика и системы управления. 2015. № 1(43). С. 93–104.
9. Левин В.А., Осипов Г.В. Управление движением группы мобильных роботов // Письма в журнал технической физики. 2016. Т. 42. Вып. 6. С. 42–48.
10. Дивеев А.И., Софронова Е.А., Шмалько Е.Ю. Эволюционные численные методы решения задачи синтеза системы управления группой роботов. // Информационные и математические технологии в науке и управлении. 2016. № 3. С. 11–22.
11. Шаповалов И.О., Косенко К.Ю. Распределенное нелинейное управление группой роботов на основе квазилинейной формы уравнений. // Известия Южного федерального университета. Технические науки. 2014. № 5. С. 45–51.
12. Зенкевич С.Л., Хуа Чжу Управление движением группы роботов в строю типа «конвой» // Мехатроника, автоматизация, управление. 2017. Т. 18. № 1. С. 30–34. DOI: 10.17587/mau.18.30-34
13. Каляев И.А. Децентрализованная система планирования и управления коллективом транспортных роботов // Кибернетика. 1985. № 4. С. 93–97.
14. Шаповалов И.О. Применение групп мобильных роботов в сложных транспортных задачах // Известия Южного федерального университета. 2012. № 2(127). С. 141–146.
15. Каляев И.А. Система планирования и управления деятельностью коллектива транспортных роботов // Методы автоматизации и проектирования, программирования и моделирования. 1982. Вып. 3. С. 119–123.
16. Закин Я.Х. Прикладная теория движения автопоезда. М.: Транспорт, 1967. 255 с.
17. Неймарк Ю.И., Фуфаев Н.А. Динамика неголономных систем. М.: Наука, 1967. 520 с.
18. Борисов А.В., Мамаев И.С., Килин А.А., Бизяев И.А. Избранные задачи неголономной механики. Институт компьютерных исследований, 2016. 883 с.
19. Зегжда С.А., Юшков М.П., Солтаханов Ш.Х. Неголономная механика. Теория и приложения. М.: Наука, Физматлит, 2009. 344 с.
20. Лурье А.И. Аналитическая механика. М.: ГИФМЛ, 1961. 824 с.
21. Добронравов В.В. Основы механики неголономных систем. М.: Высш. школа, 1970. 272 с.
22. Маркеев А.П. Теоретическая механика. М., Ижевск: Регулярная и хаотическая механика, 2007. 592 с.

Статья поступила в редакцию 02.02.2017

REFERENCES

1. **Yurevich Ye.I., Novachenko S.I., Pavlov V.A.** *Upravleniye robotami ot EVM [Control of robots from a computer]*. Moscow: Nauka Publ., 1980. 261 p. (rus)
2. **Makarov I.M., Rakhmankulov V.Z.** Gruppovoye upravleniye robotami-manipulyatorami s raspredelenno-tsentralizovannoy organizatsiyey obrabotki informatsii [Group management of robotic manipulators with distributed-centralized organization of information processing]. *Mikroprotse-sornyye sistemy upravleniya v radiotekhnike [Microprocessor control systems in radio engineering]*. Moscow: Nauka Publ., 1984, Pp. 35–45. (rus)
3. **Kalyayev I.A., Gayduk A.R., Kapustyan S.G.** *Modeli i algoritmy kollektivnogo upravleniya v gruppakh robotov [Models and algorithms of collective management in groups of robots]*. Moscow: Fizmatlit Publ., 2009, 280 p. (rus)
4. **Stroyev V.** Sistemy s iskusstvennym intellektom v sukhoputnykh voyskakh [Systems with artificial intelligence in the Army]. *Zarubezhnoye voyennoye obozreniye [Foreign military review]*, 1997, No. 3, Pp. 27–30. (rus)
5. **Rubtsov I.V., Nesterov V.Ye., Rubtsov V.I.** Sovremennaya zarubezhnaya voyennaya mikro- i mini-robototekhnika [Modern foreign military micro- and mini-robotics]. *Mikrosistemnaya tekhnika [Microsystems Technology]*, 2000, No. 3, Pp. 36–42. (rus)
6. **Kalyayev I.A., Gayduk A.R., Kapustyan S.G.** *Raspredelennyye sistemy planirovaniya deystviy kollektivov robotov [Distributed systems for planning the actions of teams of robots]*, Moscow: Yanus-K Publ., 2002, 291 p. (rus)
7. **Shiryayev V.I.** Ob upravlenii kollektivom robotov pri igre v futbole kak zadache upravleniya v usloviyakh netochnoy informatsii [On the management of a team of robots in the game of football as a management task in conditions of inaccurate information]. *Iskusstvennyy intellekt [Artificial intelligence]*, 2000, No. 3, Pp. 353–360. (rus)
8. **Shcherbatov I.A., Protalinskiy I.O., Protalinskiy O.M.** Upravleniye gruppy robotov: komponentnyy podkhod [Control of the robots group: componential approach]. *Informatika i sistemy upravleniya [Information Science and Control Systems]*, 2015, No. 1(43), Pp. 93–104. (rus)
9. **Levin V.A., Osipov G.V.** Upravleniye dvizheniyem gruppy mobilnykh robotov [Movement control of a group of mobile robots]. *Pisma v zhurnal tekhnicheskoy fiziki [Technical Physics Letters]*, 2016, Vol. 42, No. 6, Pp. 42–48. (rus)
10. **Diveyev A.I., Sofronova Ye.A., Shmalko Ye.Yu.** Evolyutsionnyye chislennyye metody resheniya zadachi sinteza sistemy upravleniya gruppy robotov [Evolutionary numerical methods for solving the problem of synthesis of a robot group control system]. *Informatsionnyye i matematicheskiye tekhnologii v nauke i upravlenii [Information and Mathematical Technologies in Science and Management]*, 2016, No. 3, Pp. 11–22. (rus)
11. **Shapovalov I.O., Kosenko K.Yu.** Raspredelennoye nelineynoye upravleniye gruppy robotov na osnove kvazilineynoy formy uravneniy [Distributed motion control system for group of large-size objects]. *Izvestiya Yuzhnogo federalnogo universiteta. Tekhnicheskkiye nauki [Izvestiya SFedU. Engineering Sciences]*, 2014, No. 5, Pp. 45–51. (rus)
12. **Zenkevich S.L., Khua Chzhu.** Upravleniye dvizheniyem gruppy robotov v stroyu tipa “konvoy” [Control of a group of mobile robots moving in the convoy type formation]. *Mekhatronika, avtomatizatsiya, upravleniya [Mechatronics, Automation, Control]*, 2017, Vol. 18, No. 1, Pp. 30–34. DOI: 10.17587/mau.18.30-34 (rus)
13. **Kalyayev I.A.** Detsentralizovannaya sistema planirovaniya i upravleniya kollektivom transportnykh robotov poborov [Decentralized system for planning and managing a team of transport robots]. *Kibernetika [Cybernetics]*, 1985, No. 4, Pp. 93–97. (rus)
14. **Shapovalov I.O.** Primeneniye grupp mobilnykh robotov v slozhnykh transportnykh zadachakh [Mobile robot group using in complex transport problems]. *Izvestiya Yuzhnogo federalnogo universiteta. Tekhnicheskkiye nauki [Izvestiya SFedU. Engineering Sciences]*, 2012, No. 2(127), Pp. 141–146. (rus)
15. **Kalyayev I.A.** Sistema planirovaniya i upravleniya deyatelnostyu kollektiva transportnykh robotov [System for planning and managing the activity of the collective of transport robots]. *Metody avtomatizatsii i proyektirovaniya, programmirovaniya i modelirovaniya [Methods of automation and design, programming and modeling]*, 1982, Vol. 3, Pp.119–123. (rus)
16. **Zakin Ya.Kh.** *Prikladnaya teoriya dvizheniya avtopoyezda [Applied Motion Theory]*. Moscow: Transport Publ., 1967, 255 p. (rus)
17. **Neymark Yu.I., Fufayev N.A.** *Dinamika negolonomnykh system [Dynamics of non-holonomic systems]*. Moscow: Nauka Publ., 1967, 520 p. (rus)
18. **Borisov A.V., Mamayev I.S., Kilin A.A., Bizyayev I.A.** *Izbrannyye zadachi negolonomnoy mekhaniki [Selected problems of non-holonomic*

mechanics]. Institut kompyuternykh issledovaniy Publ., 2016, 883 p. (rus)

19. **Zegzhda S.A., Yushkov M.P., Soltakhanov Sh.Kh.** *Negolonomnaya mekhanika. Teoriya i prilozheniya* [*Non-head mechanics. Theory and applications*]. Moscow: Nauka, Fizmatlit Publ., 2009, 344 p. (rus)

20. **Lurye A.I.** *Analiticheskaya mekhanika* [*Analytical Mechanics*]. Moscow: GIFML Publ.,

1961, 824 p. (rus)

21. **Dobronravov V.V.** *Osnovy mekhaniki negolonomnykh system* [*Fundamentals of mechanics of non-holonomic systems*]. Moscow: Vysshaya shkola Publ., 1970, 272 p. (rus)

22. **Markeyev A.P.** *Teoreticheskaya mekhanika* [*Regular and chaotic mechanics*]. Moscow, Izhevsk: Regulyarnaya i khaoticheskaya mekhanika Publ., 2007, 592 p. (rus)

Received 02.02.2017

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

СМИРНОВ Алексей Сергеевич

SMIRNOV Alexey S.

E-mail: smirnov.alexey.1994@gmail.com

СМОЛЬНИКОВ Борис Александрович

SMOLNIKOV Boris A.

E-mail: smolnikovba@yandex.ru

ЛЕОНТЬЕВ Виктор Анатольевич

LEONTEV Victor A.

E-mail: vleont@mail.ru

DOI: 10.18721/JCSTCS.10308

УДК 621.375.8, 004.9, 51.74

НОВАЯ МЕТОДИКА РАСЧЕТА ОПТИЧЕСКОГО УСИЛИТЕЛЯ EDFA С МНОГОКАНАЛЬНЫМИ ИСТОЧНИКАМИ СИГНАЛОВ И НАКАЧЕК

П.А. Чаймарданов, М.С. Былина

Санкт-Петербургский государственный университет телекоммуникаций
имени профессора М.А. Бонч-Бруевича, Санкт-Петербург, Российская Федерация

В волоконно-оптических системах передачи (ВОСП) с плотным спектральным уплотнением (DWDM) широко используются оптические усилители (ОУ) на основе активных волокон, легированных ионами эрбия (EDFA). Проектированию таких ВОСП обычно предшествует их моделирование, предполагающее наличие адекватных математических моделей всех компонентов. Моделью ОУ EDFA является система обыкновенных дифференциальных уравнений (ОДУ), решив которую, можно получить распределение мощностей сигналов, накачек и шумов усиленного спонтанного излучения вдоль активного оптического волокна (АОВ). Решение данной системы является сложной задачей, что обусловлено ее жесткостью, а также особенностями задания начальных условий, которые для различных уравнений задаются в разных точках АОВ. В статье предложена новая методика численного решения системы ОДУ, моделирующая работу ОУ EDFA, обладающая большей скоростью расчета по сравнению с известными методиками и учитывающая особенности данной системы.

Ключевые слова: оптический усилитель; EDFA; DWDM; ВОСП; моделирование.

Ссылка при цитировании: Чаймарданов П.А., Былина М.С. Новая методика расчета оптического усилителя EDFA с многоканальными источниками сигналов и накачек // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2017. Т. 10. № 3. С. 92–102. DOI: 10.18721/JCSTCS.10308

A NOVEL CALCULATION METHOD FOR OPTICAL AMPLIFIER EDFA WITH MULTICHANNEL SIGNAL AND PUMP SOURCES

P.A. Chaymardanov, M.S. Bylina

Bonch-Bruevich Saint-Petersburg State University of Telecommunications,
St. Petersburg, Russian Federation

Optical amplifier (OA) based on erbium-doped fiber (EDFA, Erbium-Doped Fiber Amplifier) are widely used in optical fiber transmission systems with dense wavelength division multiplexing (DWDM). The design of the optical fiber transmission systems is usually preceded by a simulation, assuming that there are adequate mathematical models of all components. The model of the OA EDFA is a system of ordinary differential equations (ODE) whose solution allows to obtain the power distribution of signals, pumping and noise of amplified spontaneous emission along the erbium doped fiber. The solution of this system is a complex problem, which is due to its stiffness, as well as the specifics of setting the initial conditions, which for different equations are set at different points of the erbium-doped fiber. This paper presents a novel method for the numerical solution of the ODE system, simulating the operation of the OA EDFA, which has a higher calculation speed in comparison with the known

methods and takes into account the features of this system.

Keywords: optical amplifier; EDFA; DWDM; optical fiber transmission systems; simulation.

Citation: Chaymardanov P.A, Bylina M.S. A novel calculation method for optical amplifier EDFA with multichannel signal and pump sources. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems. 2017, Vol. 10, No. 3, Pp. 92–102. DOI: 10.18721/JCSTCS.10308

В волоконно-оптических системах передачи (ВОСП) с плотным спектральным уплотнением (Dense Wavelength Division Multiplexing – DWDM) широко применяются оптические усилители (ОУ) на основе активных оптических волокон (АОВ), легированных ионами эрбия (Erbium Doped Fiber Amplifier – EDFA), способные усиливать многоканальный оптический сигнал, расположенный в диапазоне 1525–1575 нм. Важнейшим этапом проектирования ВОСП является имитационное моделирование, позволяющее проанализировать различные варианты построения ВОСП с целью оптимизации ее технико-экономических характеристик. Построение модели ВОСП предполагает предварительное моделирование всех ее компонентов.

ОУ EDFA представляет собой один из наиболее сложных для моделирования компонентов ВОСП. Моделью ОУ EDFA является система, составленная из обыкновенных дифференциальных уравнений (ОДУ), описывающих распределение мощностей сигналов, накачек и шумов усиленного спонтанного излучения (Amplified Spontaneous Emission – ASE) вдоль АОВ [1–3]. Данная система в общем случае не может быть решена аналитически, поэтому реша-

ется численно. При численном решении этой системы необходимо учитывать две ее особенности. Во-первых, она может проявлять свойство жесткости [4], что может привести к неустойчивому решению при неправильном выборе метода или шага интегрирования. Во-вторых, начальные условия для моделирования задаются в разных точках АОВ, при этом для численных методов решения систем ОДУ необходимо задание всех начальных условий только в одной точке. Для решения данных проблем необходимо использовать вспомогательные методики, позволяющие задавать начальные условия в разных точках АОВ и контролировать шаг интегрирования, что значительно увеличивает время расчета, поэтому разработка методик, увеличивающих скорость расчета ОУ EDFA в ВОСП DWDM, является актуальной задачей.

В настоящей статье приведен краткий обзор способов решения подобных систем ОДУ, предложена новая методика расчета ОУ EDFA, учитывающая все особенности данной системы ОДУ и существенно выигрывающая в скорости расчета по сравнению с известными методиками.

Структурная схема EDFA представлена на рис. 1. Она включает источник многока-

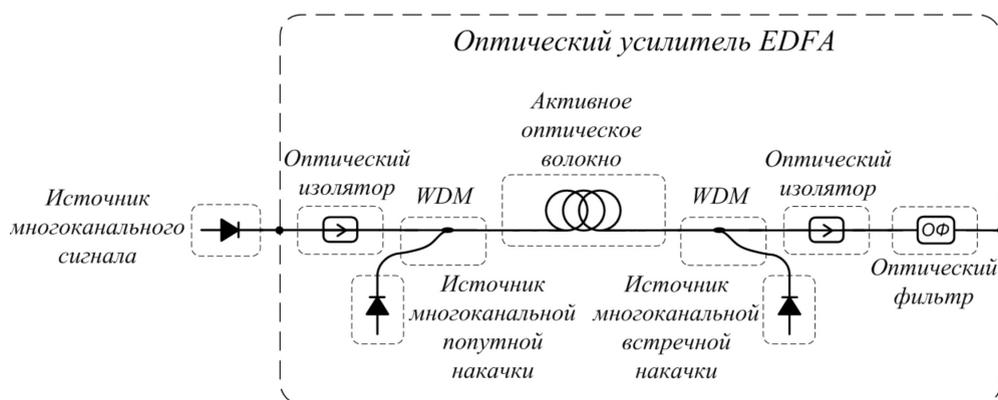


Рис. 1. Схема усилителя EDFA с многоканальными источниками сигналов и накачек

нального сигнала (ИМС), источники многоканальной попутной (ИМПН) и встречной (ИМВН) накачки, АОВ длиной L , мультиплексоры спектрального уплотнения (Wavelength Division Multiplexing – WDM), оптические изоляторы (ОИ) и оптический фильтр (ОФ).

Многоканальный сигнал с выхода источника поступает на вход усилителя. ИМПН и ИМВН создают многоканальную оптическую накачку в диапазоне длин волн вблизи 980 нм и/или 1480 нм, которая распространяется в попутном и/или встречном направлении по отношению к сигналу. Объединение на входе АОВ и разделение на его выходе сигнала и накачек осуществляется мультиплексорами WDM. Усиление сигнала происходит в АОВ. ОИ на входе ОУ EDFA предотвращает попадание шумов из усилителя в предыдущие каскады в схеме ВОСП. ОИ на выходе ОУ EDFA предотвращает попадание в усилитель шумов из последующих усилительных каскадов в схеме ВОСП. ОФ необходим для сглаживания неравномерности коэффициента усиления в полосе усиления.

Процесс усиления в АОВ описывается при помощи трехуровневой квантовой системы [1–3], в соответствии с которой ион эрбия в АОВ в каждый момент времени может находиться на одном из трех энергетических уровней. Уровень 1 с наименьшей энергией является основным, а уровень 2 – метастабильным, т. е. среднее время жизни τ иона на уровне 2 существенно больше, чем τ_3 на уровне 3. Усиление сигнала, поглощение сигнала и накачки и генерация шума ASE в АОВ происходят за счет переходов ионов эрбия между этими уровнями с поглощением или испусканием квантов света.

Рассматриваемая в данной статье модель ОУ EDFA полагается стационарной (не зависящей от времени), а также полагается, что все ионы эрбия, переходящие на уровень 3, сразу же оказываются на уровне 2, так как $\tau_3 \ll \tau$.

Модель ОУ EDFA, полученная путем обобщения моделей, представленных в [1–3], имеет вид [5]:

$$\begin{aligned} W_{12} &= \sum \frac{P_s \alpha_{12}(v_s)}{h v_s \tau \varepsilon} + \\ &+ \sum \frac{(P_{ASE}^+ + P_{ASE}^-) \alpha_{12}(v_{ASE})}{h v_{ASE} \tau \varepsilon} \\ W_{21} &= \sum \frac{P_s g_{21}(v_s)}{h v_s \tau \varepsilon} + \\ &+ \sum \frac{(P_{ASE}^+ + P_{ASE}^-) g_{21}(v_{ASE})}{h v_{ASE} \tau \varepsilon} \\ R_{12} &= \sum \frac{(P_{p1480}^+ + P_{p1480}^-) \alpha_{12}(v_{p1480})}{h v_{p1480} \tau \varepsilon} \\ R_{21} &= \sum \frac{(P_{p1480}^+ + P_{p1480}^-) g_{12}(v_{p1480})}{h v_{p1480} \tau \varepsilon} \\ R_{13} &= \sum \frac{(P_{p980}^+ + P_{p980}^-) \alpha_{13}(v_{p980})}{h v_{p980} \tau \varepsilon} \\ \varepsilon &= \frac{\pi b_{eff}^2 N}{\tau} \end{aligned} \quad (1)$$

$$\begin{aligned} N_{2,отн} &= (W_{12} + R_{12} + R_{13}) / \\ &/ (W_{12} + R_{12} + W_{21} + R_{21} + R_{13} + 1/\tau) \\ dP_{p1480}^{\pm} / dz &= \mp P_{p1480}^{\pm} [\alpha_{12}(v_{p1480})(1 - N_{2,отн}) - \\ &- g_{21}(v_{p1480})N_{2,отн}] \mp \alpha_{p1480}(v_{p1480})P_{p1480}^{\pm} \\ dP_{p980}^{\pm} / dz &= \mp P_{p980}^{\pm} [\alpha_{13}(v_{p980})(1 - N_{2,отн})] \mp \\ &\mp \alpha_{p980}(v_{p980})P_{p980}^{\pm} \\ dP_s / dz &= P_s [g_{21}(v_s)N_{2,отн} - \\ &- \alpha_{12}(v_s)(1 - N_{2,отн})] - \alpha_s(v_s)P_s \\ dP_{ASE}^{\pm} / dz &= \pm P_{ASE}^{\pm} [g_{21}(v_{ASE})N_{2,отн} - \\ &- \alpha_{12}(v_{ASE})(1 - N_{2,отн})] \pm 2g_{21}(v_{ASE}) \times \\ &\times N_{2,отн} h v_{ASE} \Delta v \mp \alpha_{ASE}(v_{ASE})P_{ASE}^{\pm}, \end{aligned}$$

где суммирование производится по количеству имеющихся каналов для сигналов, накачек и шумов ASE. Отметим, что спектр шумов ASE непрерывен и существует во всей полосе усиления АОВ, поэтому в модели предусмотрена дискретизация спектра шумов ASE на несколько шумовых каналов с небольшой (по сравнению с шириной полосы усиления АОВ) полосой Δv .

В (1) использованы следующие обозначения: R_{ij} и W_{ij} – вероятности вынужденных переходов ионов эрбия с уровня i на

уровень j , вызванные поглощением фотонов накачки, сигналов и шумов ASE; N – концентрация ионов эрбия; $N_{2,отн}$ – относительная концентрация ионов эрбия на уровне 2 (по отношению к общей концентрации ионов эрбия); $\nu_s, \nu_{ASE}, \nu_{p1480}, \nu_{p980}$ – центральные частоты каналов сигнала, ASE, накачек вблизи длин волн 1480 и 980 нм; α_{12}, g_{21} – коэффициенты поглощения и усиления в полосе сигнала и накачки 1480 нм, зависящие от температуры АОВ и частоты канала; α_{13} – коэффициент поглощения в полосе накачки 980 нм, зависящий от температуры АОВ и частоты канала; $\alpha_s, \alpha_{ASE}, \alpha_{p1480}, \alpha_{p980}$ – коэффициенты, учитывающие дополнительные потери сигнала, ASE, накачек вблизи длин волн 1480 и 980 нм в АОВ, не вызванные поглощением ионами эрбия; b_{eff} – эффективный радиус распределения ионов эрбия; ε – параметр насыщения; P_s – мощность сигнала в определенном канале; $P_{p1480}^+, P_{p980}^+, P_{ASE}^+$ – мощности накачек на длинах волн 1480 и 980 нм и шумов ASE в попутном (+) и встречном (–) направлениях в определенных каналах; h – постоянная Планка; z – координата вдоль АОВ.

Поскольку параметры АОВ зависят от температуры T [6], при моделировании необходимо учесть данную зависимость при помощи формул [7]:

$$\begin{aligned} \alpha_{12}(\lambda, T) &= \alpha_{12}(\lambda, \infty) \exp\left(\frac{\beta_{12}(\lambda)}{KT}\right), \\ \alpha_{13}(\lambda, T) &= \alpha_{13}(\lambda, \infty) \exp\left(\frac{\beta_{13}(\lambda)}{KT}\right), \\ g_{21}(\lambda, T) &= g_{21}(\lambda, \infty) \exp\left(\frac{\beta_{21}(\lambda)}{KT}\right), \end{aligned} \quad (2)$$

где K – постоянная Больцмана; λ – длина волны; $\alpha_{12}(\lambda, \infty)$ и $\beta_{12}(\lambda)$, $\alpha_{13}(\lambda, \infty)$ и $\beta_{13}(\lambda)$, $g_{21}(\lambda, \infty)$ и $\beta_{21}(\lambda)$ – коэффициенты, которые можно определить по известным зависимостям соответствующих коэффициентов поглощения и усиления от длины волны при двух разных температурах.

Система уравнений (1) с учетом (2) для рассматриваемого случая, представляет собой наиболее распространенную в литературе модель ОУ EDFA в стационар-

ном режиме. В общем случае она не имеет аналитического решения, поэтому должна решаться численно. Однако получение численного решения может оказаться затруднительным, что обусловлено двумя отмеченными выше особенностями данной системы.

Рассмотрим подробнее первую особенность системы (1). Уравнения распределения мощностей сигнала, накачек и шумов ASE по длине АОВ (1) представляют собой задачу Коши для системы ОДУ первого порядка с заданным начальным условием в точке z_0 или z_L :

$$\begin{aligned} P' &= f(z, P), \\ P(z_0) &= P_0, \\ P(z_L) &= P_L, \\ z_0 &\leq z \leq z_L, \end{aligned} \quad (3)$$

где P – n -мерный вектор переменных состояния (мощностей); P_0, P_L – n -мерные векторы начальных и конечных данных задачи Коши (значений мощностей в начале и в конце АОВ); f – n -мерный вектор-функция правых частей уравнений (1); z – независимая переменная (расстояние вдоль АОВ); n – количество уравнений (1).

Как было сказано ранее, система (1) может проявлять свойства жесткости. Общепринятого определения жесткости системы не существует, но в ряде источников [4, 8] приводится описание жесткой задачи. В [4, 8] систему ОДУ называют жесткой, если хотя бы одно значение в матрице Якоби, составленной из $J_{ij} = \partial f_i / \partial P_j$ (i и j – номера уравнений в системе ОДУ), для уравнений системы (3), на некотором интервале $z_0 \leq z \leq z_L$, отрицательно, или $s = \max(|\operatorname{Re}(J_{ij})|) / \min(|\operatorname{Re}(J_{ij})|) \gg 1$, где s называют коэффициентом жесткости данной системы. Системы с $s > 10$, согласно [8], можно считать жесткими.

Исследование системы (1) для ОУ с различными конфигурациями накачек показало, что она может проявлять свойство жесткости, при этом увеличение коэффициента жесткости обусловлено большой разницей собственных значений матрицы Якоби, возникающей вследствие различных скоростей изменения мощностей накачек и

сигналов вдоль АОВ.

Жесткость системы (1) может привести к некорректным результатам моделирования при использовании явных численных методов решения систем ОДУ, таких как методы Эйлера, Рунге-Кутты 3, 4 порядков и др. При использовании явных методов необходимо неоправданно уменьшать шаг интегрирования [9], чтобы получить устойчивое решение.

Для решения жестких систем ОДУ обычно используют [4] неявные методы численного интегрирования, такие как методы Гира различных порядков [8, 10, 11], Розенброка [12], неявные методы Рунге-Кутта [13–15] и др. Данные методы обладают лучшей, по сравнению с явными схемами решения, устойчивостью и не накладывают серьезных ограничений на шаг интегрирования. Однако для обеспечения устойчивости этих методов шаг все равно требуется контролировать [8].

Принимая во внимание сказанное выше, для учета первой особенности системы (1) необходимо при выборе метода интегрирования опираться на неявные схемы решения ОДУ, а также контролировать шаг интегрирования для обеспечения устойчивости метода.

Вторая особенность системы (1) связана с заданием начальных условий в разных точках АОВ. Для численных методов интегрирования необходимо задание начальных условий в одной начальной точке для всех уравнений. Для решения данной задачи используются методы стрельбы, релаксационные методы [16], генетические алгоритмы [17].

Метод стрельбы подразумевает подбор таких значений $P_{p1480}^-(0)$, $P_{p980}^-(0)$ и $P_{ASE}^-(0)$, при которых после решения системы (1) полученные значения $P_{p1480}^-(L)$, $P_{p980}^-(L)$ и $P_{ASE}^-(L)$ совпали бы с известными в пределах заданной погрешности. Основным недостатком этого метода является большое время расчета. Улучшенным вариантом метода стрельбы для нескольких функций является метод Ньютона–Рафсона [18], позволяющий производить пристрелку одновременно для нескольких уравнений в системе ОДУ, однако данный метод также

требует большого времени расчета.

Релаксационный метод также подразумевает подбор значений $P_{p1480}^-(0)$, $P_{p980}^-(0)$ и $P_{ASE}^-(0)$ с использованием следующего алгоритма:

1. Задание в качестве начальных условий $P_{p1480}^-(0) = 0$, $P_{p980}^-(0) = 0$, $P_{ASE}^-(0) = 0$.

2. Решение системы (1) при изменении z от 0 до L . Завершение расчета, если полученные значения $P_{p1480}^-(L)$, $P_{p980}^-(L)$ и $P_{ASE}^-(L)$ совпали с известными в пределах заданной погрешности.

3. Задание в качестве начальных условий для этапа 4 значений $P_{p1480}^-(L)$, $P_{p980}^-(L)$ и $P_{ASE}^-(L)$, полученных на этапе 2.

4. Решение системы (1) при изменении z от L до 0.

5. Задание в качестве начальных условий для этапа 2 значений $P_{p1480}^-(0)$, $P_{p980}^-(0)$ и $P_{ASE}^-(0)$, полученных на этапе 4. Возвращение к этапу 2.

Недостатками данного метода являются неустойчивость решения при некоторых конфигурациях усилителя, а также большое время расчета.

Методика на основе генетического алгоритма (ГА) также является улучшенной версией метода стрельбы. В рамках данной методики при помощи ГА производится подбор недостающих условий в точке $z = 0$ и решается система уравнений (1). Полный текст методики представлен в [17]. По сравнению с методом стрельбы и релаксационным методом данная методика имеет меньшее, но все же существенное время расчета.

Учитывая потенциальную неустойчивость решения системы (1) численными методами и трудности с выбором начальных условий, была разработана новая улучшенная методика расчета ОУ EDFA.

Суть предлагаемой методики заключается в следующем. Предположим, что система (1) решена при заданных начальных условиях. Тогда АОВ характеризуется функцией $N_{2,отн}(z)$, однозначно соответствующей моделируемому случаю. Целью разработанной методики является нахождение $N_{2,отн}(z)$ в каждой точке АОВ. Методика подразумевает получение этой функции итеративным методом, путем независимого решения в

пределах i -й итерации уравнений только для попутных каналов и только встречных каналов.

Заменим уравнения в системе (1) с учетом обозначенного предположения на две группы уравнений для каналов в попутном (4) и встречном (5) направлениях (зависимости коэффициентов поглощения, усиления и дополнительных потерь от центральных частот каналов предполагаются, но в представленных формулах будут опущены).

Для попутных каналов:

$$\begin{aligned}
 P_s(z) &= P_s(0) \exp((\alpha_{12} + g_{21}) \times \\
 &\times \int_0^z N_{2.отн}(\tau) d\tau - \alpha_{12}z - \alpha_s z), \\
 P_{p1480}^+(z) &= P_{p1480}^+(0) \exp((\alpha_{12} + g_{21}) \times \\
 &\times \int_0^z N_{2.отн}(\tau) d\tau - \alpha_{12}z - \alpha_{p1480}z), \\
 P_{p980}^+(z) &= P_{p980}^+(0) \exp(\alpha_{13} \int_0^z N_{2.отн}(\tau) d\tau - \\
 &- \alpha_{13}z - \alpha_{p980}z), \\
 P_{ASE}^+(z) &= (P_{ASE}^+(0) + \int_0^z 2g_{21}h\nu_{ASE}\Delta\nu N_{2.отн}(\tau) \times \\
 &\times \exp(-(\alpha_{12} + g_{21}) \int_0^\tau N_{2.отн}(\sigma) d\sigma + \\
 &+ \alpha_{12}\tau + \alpha_{ASE}\tau) d\tau) \exp((\alpha_{12} + g_{21}) \times \\
 &\times \int_0^z N_{2.отн}(\tau) d\tau - \alpha_{12}z - \alpha_{ASE}z).
 \end{aligned}
 \tag{4}$$

Для встречных каналов:

$$\begin{aligned}
 P_{p1480}^-(z) &= P_{p1480}^-(L) \exp((\alpha_{12} + g_{21}) \times \\
 &\times \int_0^{L-z} N_{2.отн}(L-\tau) d\tau - \alpha_{12}(L-z) - \\
 &- \alpha_{p1480}(L-z)), \\
 P_{p980}^-(z) &= P_{p980}^-(L) \exp(\alpha_{13} \int_0^{L-z} N_{2.отн}(L-\tau) d\tau - \\
 &- \alpha_{13}(L-z) - \alpha_{p980}(L-z)), \\
 P_{ASE}^-(z) &= (P_{ASE}^-(L) +
 \end{aligned}
 \tag{5}$$

$$\begin{aligned}
 &+ \int_0^{L-z} 2g_{21}h\nu_{ASE}\Delta\nu N_{2.отн}(L-\tau) \times \\
 &\times \exp(-(\alpha_{12} + g_{21}) \int_0^\tau N_{2.отн}(L-\sigma) d\sigma + \\
 &+ \alpha_{12}(L-\tau) + \alpha_{ASE}(L-\tau)) d\tau) \times \\
 &\times \exp((\alpha_{12} + g_{21}) \int_0^{L-z} N_{2.отн}(L-\tau) d\tau - \\
 &- \alpha_{12}(L-z) - \alpha_{ASE}(L-z)).
 \end{aligned}
 \tag{5}$$

Разобьем всю длину АОВ на l равноотстоящих точек с шагом Δz . Выразим значения функций (4) и (5) в точках $k+1$ и $k-1$ через их значения в точке k при помощи метода прямоугольников.

Для попутных каналов:

$$\begin{aligned}
 P_{s,i,k+1} &\approx P_{s,i,k} A^{N_{2.отн,i,k}} B, \\
 P_{p1480,i,k+1}^+ &\approx P_{p1480,i,k}^+ A^{N_{2.отн,i,k}} B, \\
 P_{p980,i,k+1}^+ &\approx P_{p980,i,k}^+ C^{N_{2.отн,i,k}} D, \\
 P_{ASE,i,k+1}^+ &\approx (P_{ASE,i,k}^+ + N_{2.отн,i,k} E \times \\
 &\times (A^{N_{2.отн,i,k}} B)^{-1}) A^{N_{2.отн,i,k}} B.
 \end{aligned}
 \tag{6}$$

Для встречных каналов:

$$\begin{aligned}
 P_{1480,i,k-1}^- &\approx P_{1480,i,k}^- A^{N_{2.отн,i,k}} B, \\
 P_{980,i,k-1}^- &\approx P_{980,i,k}^- C^{N_{2.отн,i,k}} D, \\
 P_{ASE,i,k-1}^- &\approx (P_{ASE,i,k}^- + N_{2.отн,i,k} E \times \\
 &\times (A^{N_{2.отн,i,k}} B)^{-1}) A^{N_{2.отн,i,k}} B,
 \end{aligned}
 \tag{7}$$

где A, B, C, D, E – константы (вычисляемые для каждого канала (6) и (7)), для которых справедливо:

$$\begin{aligned}
 A &= \exp((\alpha_{12} + g_{21})\Delta z), \\
 B &= \exp(-(\alpha_{12} - \alpha_{доп})\Delta z), \\
 C &= \exp(\alpha_{13}\Delta z), \\
 D &= \exp(-(\alpha_{13} - \alpha_{доп})\Delta z), \\
 E &= 2h\nu\Delta\nu\Delta z.
 \end{aligned}
 \tag{8}$$

Опишем методику расчета с учетом (6)–(8).

1. Для первой итерации ($i = 0$) полагаем, что $N_{2.отн,i-1,k} = 0$ при всех k , а мощности в каждой точке волокна для всех сигналов, накачек и шумов ASE равными $P_{i-1,k} = 0$.
2. Задаем значения $P_{s,i,0}$, $P_{p1480,i,0}^+$, $P_{p980,i,0}^+$ и $P_{ASE,i,0}^+$.
3. Задаем значения $P_{p1480,i,l}^-$, $P_{p980,i,l}^-$ и $P_{ASE,i,l}^-$.
4. Задаем ε и i_{\max} , где ε – относительная локальная погрешность расчета в каждой

точке АОВ, а i_{\max} – максимальное количество итераций.

5. Рассчитываем константы A, B, C, D, E для каждого канала по выражениям (8).

6. Производим расчет в попутном направлении. Для каждого $k \neq 0$ находим $P_{s,i,k}$, $P_{p1480,i,k}^+$, $P_{p980,i,k}^+$ и $P_{ASE,i,k}^+$ из уравнений (6), полагая $N_{2.отн,i,k} = 0,5(N_{2.отн,k} + N_{2.отн,i-1,k})$, где $N_{2.отн,k}$ рассчитывается по (1) с учетом $P_{p1480,i-1,k}^-$, $P_{p980,i-1,k}^-$ и $P_{ASE,i-1,k}^-$.

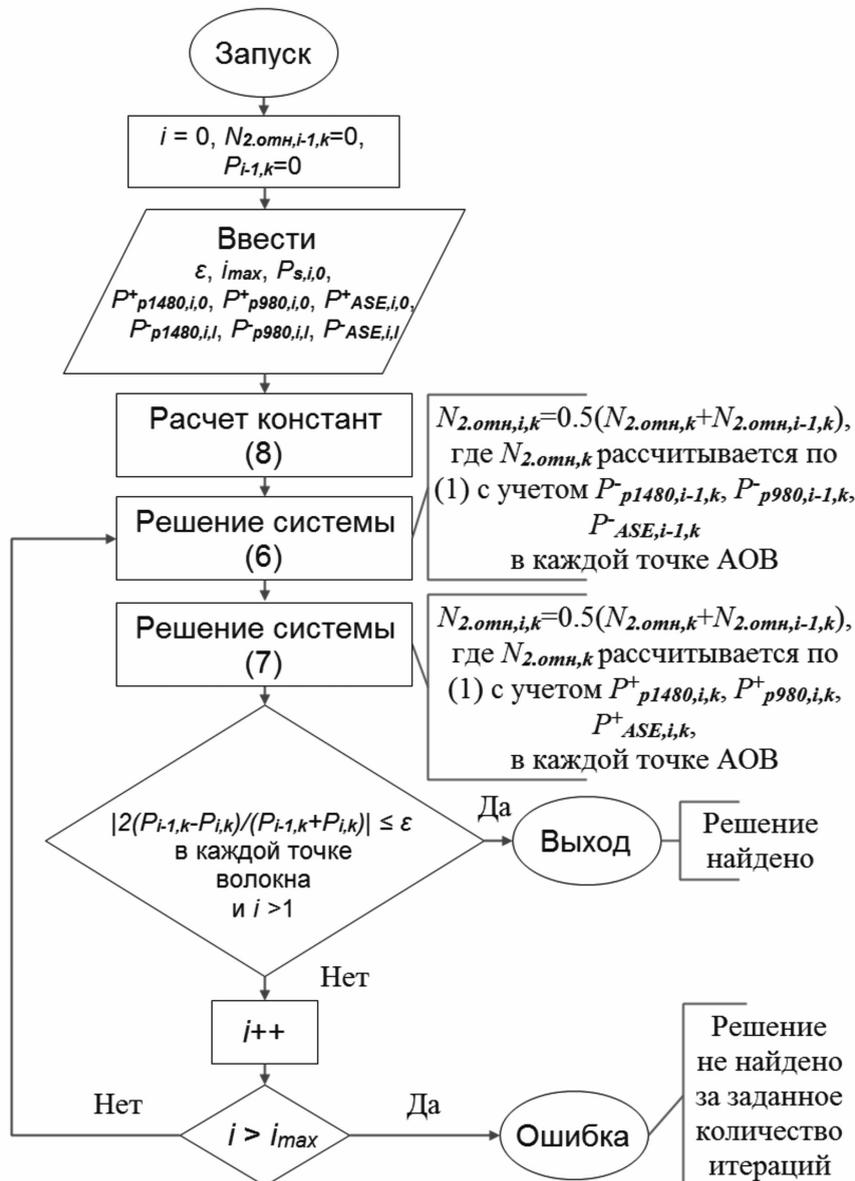


Рис. 2. Блок-схема предлагаемого алгоритма расчета

7. Производим расчет во встречном направлении. Для каждого $k \neq 1$ находим $P_{p1480,i,k}^-$, $P_{p980,i,k}^-$ и $P_{ASE,i,k}^-$ из уравнения (7), полагая $N_{2.отн,i,k} = 0,5(N_{2.отн,k} + N_{2.отн,i-1,k})$, где $N_{2.отн,k}$ рассчитывается по (1) с учетом $P_{s,i,k}$, $P_{p1480,i,k}^+$, $P_{p980,i,k}^+$ и $P_{ASE,i,k}^+$.

8. Если $i > 1$ и для всех P выполняются неравенства

$|2(P_{i-1,k} - P_{i,k}) / (P_{i-1,k} + P_{i,k})| \leq \epsilon$, то расчет заканчивается. Решением (1) считаются $N_{2.отн,i,k}$, $P_{s,i,k}$, $P_{p1480,i,k}^+$, $P_{p980,i,k}^+$, $P_{ASE,i,k}^+$, $P_{p1480,i,k}^-$, $P_{p980,i,k}^-$ и $P_{ASE,i,k}^-$.

9. Сохраняем $N_{2.отн,i-1,k} = N_{2.отн,i,k}$.

10. Увеличиваем i на 1. Если $i \leq i_{max}$, возвращаемся к п. 6. Если $i > i_{max}$, алгоритм расчета завершается с ошибкой: решение не найдено за заданное количество итераций.

Описанный алгоритм представлен в виде блок-схемы на рис. 2.

Полученные при решении (1) значения мощностей сигнала $P_{s,i,k}$ позволяют рассчитать коэффициенты усиления G для каждого сигнала в каждом канале:

$$G = 10 \log_{10}(P_{out} / P_{in}) = 10 \log_{10}(P_{s,i,l} / P_{s,i,0}) - 2a_{ОИ} - a_{WDM} - a_{ОФ}, \quad (9)$$

где P_{in} , P_{out} — мощности сигнала в канале на входе и выходе ОУ EDFA, $a_{ОИ}$, a_{WDM} , $a_{ОФ}$ — потери, вносимые ОИ, мультиплексорами WDM и ОФ соответственно.

Параметры моделируемого ОУ EDFA

Параметр	Значение
Марка активного волокна	М-5
Параметр насыщения активного волокна, 1/(м·с)	$3.047 \cdot 10^{15}$
Температура, °С	25
Длина активного волокна, м	15
Вид накачки	Встречная
Число каналов накачки	1
Длина волны накачки, нм	980
Мощность накачки, мВт	100

Отметим, что соотношения (6)–(8) были записаны для постоянного шага интегрирования Δz , хотя предлагаемая методика легко обобщается на случай переменного шага. Выбор фиксированного Δz позволяет сократить объем расчетов за счет ввода констант (8), которые уменьшают количество необходимых вычислений.

Для проверки правильности предложенной методики проведено моделирование ОУ EDFA с многоканальным (40 каналов) входным сигналом с равномерным расположением каналов в диапазоне 1500–1600 нм и мощностью 1 мкВт/канал. Потери, вносимые входным и выходным ОИ, мультиплексорами WDM, ОФ, принимались равными нулю. Расчет проведен при параметрах ОУ EDFA, приведенных в таблице.

Аналогичные расчеты были проведены

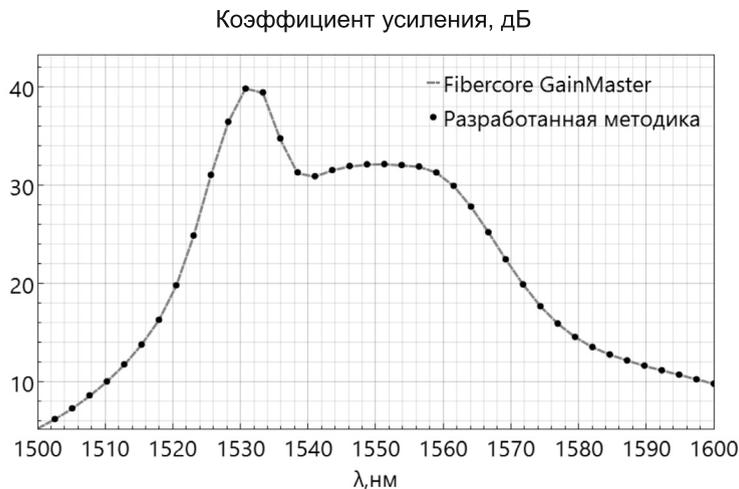


Рис. 3. Результаты сопоставления рассчитанных коэффициентов усиления

с помощью свободно распространяемой программы **Fibercore GainMaster**, моделирующей работу ОУ EDFA. Автор данной программы – производитель эрбиевых волокон для ОУ EDFA и волоконных лазеров. Данная программа предназначена исключительно для моделирования ОУ на основе волокон производства **Fibercore**. Сравнивался коэффициент усиления ОУ EDFA, полученный в обоих случаях. Из рис. 3 видно, что результаты моделирования хорошо совпадают, что подтверждает справедливость предложенной методики.

Предлагаемая методика сравнивалась по времени вычисления с методикой на базе ГА, которая является самой быстрой по сравнению с другими описанными ранее методиками. Обе методики были применены к решению (1) для 40 канальной системы в С-диапазоне. На компьютере с тактовой частотой процессора 2,4 ГГц методика на базе ГА позволила найти решение за 1325 с, а предлагаемая методика – всего за 3 с.

В заключение сформулируем краткие выводы по результатам исследования и возможностям их практического использования.

В рамках данной статьи рассмотрены известные методы и методики решения, применяемые для моделей ОУ EDFA, выделены их основные недостатки, предложена новая методика расчета ОУ EDFA, произведено сопоставление скорости расчета по предложенной методике и ранее рассмотренных методов, доказавшее ее эффективность.

На основе модели (1) и предложенной методики расчета разработана программа, позволяющая рассчитывать мощности сигналов в отдельных каналах DWDM, накачек и шумов ASE в любом сечении АОВ, определять коэффициент усиления ОУ EDFA и отношение сигнала к шуму в каждом канале. Программа может использоваться разработчиками ОУ EDFA для проверки справедливости принимаемых проектных решений.

СПИСОК ЛИТЕРАТУРЫ

1. **Becker P.C., Olsson N.A., Simpson J.R.** Erbium-Doped fiber Amplifiers. Fundamentals and Technology. Academic Press, 1999. 460 p.
2. **Desurvire E.** Erbium-Doped fiber amplifiers. Principles and applications. New York: John Wiley & Sons, 1994. 800 p.
3. **Dignonnet M.J.F.** Rare-Earth-Doped Fiber Lasers and Amplifiers. 2nd ed. Marcel Dekker, 2001. 795 p.
4. **Новиков Е.А., Шорников Ю.В.** Компьютерное моделирование жестких гибридных систем. Новосибирск: Изд-во НГТУ, 2012. 451 с.
5. **Былина М.С., Чаймарданов П.А.** Математическая модель однокаскадного EDFA усилителя с многоканальными источниками сигнала и накачки // Актуальные проблемы инфотелекоммуникаций в науке и образовании. 2016. Т. 1. № 5. С. 265–270.
6. **Francisco J. Madruga, M. Angeles Quintela, Carlos Galindez, Mauro Lomer, Jose M. Lopez Higuera.** Effects of temperature on high concentration erbium-doped fiber intrinsic parameters // Proc. SPIE 6619, 3rd European Workshop on Optical Fibre Sensors. 2007. DOI: 10.1117/12.738365
7. **Hee Gap Park, Seung Chul Yun, Young Jun Jin.** Er-doped superfluorescent fiber source with thermally stable mean wavelength // Journal of the Optical Society of Korea. 2009. Vol. 13. No. 2. Pp. 240–244. DOI: 0.3807/JOSK.2009.13.2.240
8. **Семенов М.Е., Колупаева С.Н.** Анализ областей абсолютной устойчивости неявных методов решения систем обыкновенных дифференциальных уравнений // Известия Томского политехнического университета. 2010. Т. 317. № 2. С. 16–22.
9. **Лебедев В.И.** Явные разностные схемы для решения жестких задач с комплексным или разделимым спектром // Вычислительная математика и математическая физика. 2000. Т. 40. № 12. С. 1801–1812.
10. **Турсунов Д.А., Семенов М.Е.** Построение одношагового девятиточечного блочного метода для решения жестких систем обыкновенных дифференциальных уравнений // Современные проблемы науки и образования. 2013. № 6. С. 880–880.
11. **Akinfenwa O.A., Jator S.N., Yao N.M.** An eighth order backward differentiation formula with continuous coefficients for stiff ordinary differential equations // International Journal of Mathematical and Computer Sciences. 2011. Vol. 7. No. 4. Pp. 171–176.
12. **Розенброк Х., Стори С.** Вычислительные методы для инженеров-химиков. М.: Мир, 1968. 444 с.
13. **Скворцов Л.М.** Эффективная реализация

неявных методов Рунге-Кутты второго порядка // Математическое моделирование. 2013. Т. 25. № 5. С. 15–28. DOI: 10.1134/2FS2070048213060124

14. **Julien Alexandre dit Sandretto, Alexandre Chapoutot.** Validated explicit and implicit runge-kutta methods // *Reliable Computing*. 2016. Vol. 25. Pp. 79–103.

15. **Kennedy C.A., Carpenter M.H.** Diagonally implicit Runge-Kutta methods for Ordinary Differentia Equations // *A Review*. NASA Report. Langley research center. Hampton VA 23681, 2016. 162 p.

16. **Ufuk Parali, Cahit Pınar.** Numerical analysis

Статья поступила в редакцию 14.06.2017

REFERENCES

1. **Becker P.C., Olsson N.A., Simpson J.R.** *Erbium-Doped fiber amplifiers. Fundamentals and Technology*. Academic Press, 1999, 460 p.

2. **Desurvire E.** *Erbium-Doped fiber amplifiers. Principles and applications*. John Wiley & Sons, New York, 1994, 800 p.

3. **Digonnet M.J.F.** *Rare-Earth-Doped Fiber Lasers and Amplifiers*, 2nd ed., Marcel Dekker, 2001, 795 p.

4. **Novikov Ye.A., Shornikov Yu.V.** *Kompyuternoye modelirovaniye zhestkikh gibridnykh sistem [Computer simulation of stiff hybrid systems]*. Novosibirsk: NGTU Publ., 2012, 451 p.

5. **Bylina M.S., Chaymardanov P.A.** Matematicheskaya model odnokaskadnogo EDFA usilitelya s mnogokanalnymi istochnikami signala i nakachki [Mathematical Model of a Single-Stage EDFA Amplifier with Multi-Channel Sources of Signal and Pump]. *Aktualnyye problemy infotelekomunikatsiy v nauke i obrazovanii [Actual Problems of Infotelecommunications in Science and Education]*, 2016, Vol. 1, No. 5, Pp. 265–270. (rus)

6. **Francisco J. Madruga, M. Angeles Quintela, Carlos Galindez, Mauro Lomer, Jose M. Lopez Higuera.** Effects of temperature on high concentration erbium-doped fiber intrinsic parameters. *Proc. SPIE 6619, 3rd European Workshop on Optical Fibre Sensors*, 66190W, 2007. DOI: 10.1117/12.738365.

7. **Hee Gap Park, Seung Chul Yun, Young Jun Jin.** Er-doped superfluorescent fiber source with thermally stable mean wavelength. *Journal of the Optical Society of Korea*, 2009, Vol. 13, No. 2, Pp. 240–244. DOI: 0.3807/JOSK.2009.13.2.240.

8. **Semenov M.Ye., Kolupayeva S.N.** Analiz oblastey absolyutnoy ustoychivosti neyavnykh metodov resheniya sistem obyknovennykh differentsialnykh uravneniy [Analysis of areas of absolute stability of implicit methods for solving

of gain and noise figure in different configurations of single-pass and double-pass EDFA // *Proc. SPIE 9886, Micro-Structured and Specialty Optical Fibres IV*, 98861M. 2016. DOI: 10.1117/12.2228941

17. **Priye V., Singh D.K., Arya S.C.** Novel numerical method to model multichannel Erbium-doped Fiber Amplifier // *Fiber and Integrated Optics*. 2006. Vol. 25. Pp. 375–385. DOI: 10.1080/01468030600817134

18. **Liu X.** Powerful solution for simulating nonlinear coupled equations describing bidirectionally pumped broadband Raman amplifiers // *Optics express*. 2004. Vol. 12. No. 4. Pp. 545–550.

systems of ordinary differential equations]. *Izvestiya Tomskogo politekhnicheskogo universiteta [Bulletin of the Tomsk Polytechnic University]*, 2010, Vol. 317, No. 2, Pp. 16–22. (rus)

9. **Lebedev V.I.** Yavnyye raznostnyye skhemy dlya resheniya zhestkikh zadach s kompleksnym ili razdelimym spektrom [Explicit difference schemes for solving stiff problems with a complex or separable spectrum]. *Zhurnal vychislitel'naya matematika i matematicheskaya fizika [Computational Mathematics and Mathematical Physics]*, 2000, Vol. 40, No. 12, Pp. 1801–1812. (rus)

10. **Tursunov D.A., Semenov M.Ye.** Postroyeniye odnoshagovogo devyatiochechnogo blochnogo metoda dlya resheniya zhestkikh sistem obyknovennykh differentsialnykh uravneniy [Construction one-step nine points block method for solving stiff system of ordinary differential equations]. *Sovremennyye problemy nauki i obrazovaniya [Modern problems of science and education]*, 2013, No. 6, Pp. 880–880. (rus)

11. **Akinfenwa O.A., Jator S.N., Yao N.M.** An eighth order backward differentiation formula with continuous coefficients for stiff ordinary differential equations. *International Journal of Mathematical and Computer Sciences*, 2011, Vol. 7, No. 4, Pp. 171–176.

12. **Rozenbrok Kh., Stori S.** *Vychislitelnyye metody dlya inzhenerov-khimikov [Computational methods for chemical engineers]*. Moscow: Mir Publ., 1968, 444 p. (rus)

13. **Skvortsov L.M.** Effektivnaya realizatsiya neyavnykh metodov Runge-Kutta vtorogo poryadka [Efficient implementation of second order implicit Runge-Kutta methods]. *Matematicheskoye modelirovaniye [Mathematical Models and Computer Simulations]*, 2013, Vol. 25, No. 5, Pp. 15–28. DOI: 10.1134/2FS2070048213060124 (rus)

14. **Julien Alexandre dit Sandretto, Alexandre**

Chapoutot. Validated explicit and implicit runge-kutta methods. *Reliable Computing*, 2016, Vol. 25, Pp. 79–103.

15. **Kennedy C.A., Carpenter M.H.** Diagonally implicit Runge-Kutta methods for Ordinary Differential Equations. *A Review. NASA Report*, Langley research center, Hampton VA 23681, 2016, 162 p.

16. **Ufuk Parali, Cahit Pınar.** Numerical analysis of gain and noise figure in different configurations of single-pass and double-pass EDFA. *Proc. SPIE*

Received 14.06.2017

9886, Micro-Structured and specialty optical fibres IV, 98861M, 2016. DOI: 10.1117/12.2228941.

17. **Priye V., Singh D.K., Arya S.C.** Novel numerical method to model multichannel Erbium-doped Fiber Amplifier. *Fiber and integrated optics*, 2006, Vol. 25, Pp. 375–385. DOI: 10.1080/01468030600817134.

18. **Liu X.** Powerful solution for simulating nonlinear coupled equations describing bidirectionally pumped broadband Raman amplifiers. *Optics express*, 2004, Vol. 12, No. 4, Pp. 545–550.

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

ЧАЙМАРДАНОВ Павел Александрович

CHAYMARDANOV Pavel A.

E-mail: pchai@yandex.ru

БЫЛИНА Мария Сергеевна

BYLINA Maria S.

E-mail: BylinaMaria@mail.ru



DOI: 10.18721/JCSTCS.10309

УДК 519.876.5

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛЕЙ ЭПИДЕМИИ И КЛЕТОЧНОГО АВТОМАТА ПРИ МОДЕЛИРОВАНИИ РАСПРОСТРАНЕНИЯ ИНФОРМАЦИИ В СОЦИАЛЬНЫХ СЕТЯХ

*Д.К. Горковенко*Байкальский государственный университет,
г. Иркутск, Российская Федерация

Обоснована актуальность исследования социальных сетей. Получены данные о записях в социальных сетях и количестве просмотров в течение времени. Построен процесс оценки параметров моделей SIR и клеточного автомата. Для оценки параметров применен эволюционный генетический алгоритм. Рассмотрено несколько видов моделей эпидемии и клеточного автомата для описания процесса распространения информации. В отличие от существующих исследований впервые проведены обзор и сравнение обоих типов моделей. Проведено имитационное моделирование процесса распространения информации в социальных сетях. В ходе исследования созданы графики, в которых отражены результаты моделирования. Представлен сравнительный анализ моделей, полученных в процессе подбора параметров. В результате выбрана модель МАПП, как наиболее подходящая при описании процесса распространения информации.

Ключевые слова: социальные сети; математическое моделирование; распространение информации в сети; клеточный автомат; модель адаптивно-подражательного поведения.

Ссылка при цитировании: Горковенко Д.К. Сравнительный анализ моделей эпидемии и клеточного автомата при моделировании распространения информации в социальных сетях // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2017. Т. 10. № 3. С. 103–113. DOI: 10.18721/JCSTCS.10309

COMPARISON OF EPIDEMIC MODELS AND CELLULAR AUTOMATA IN MODELING OF DIFFUSION OF INFORMATION IN SOCIAL

D.K. Gorkovenko

Baikal State University, Irkutsk, Russian Federation

The article substantiates the importance of social network research. We have obtained data about the records in social networks and the number of record views. The process of estimating the parameters of the SIR models and the cellular automaton is constructed. The article considers several types of epidemic models and a cellular automaton for describing the process of information dissemination. We have carried out simulation modeling of the process of information dissemination in social networks. In the course of the study, the graphs reflecting the results of modeling were created. As a result, we have presented the best variant from the models. It is a model of adaptive-imitative behavior that is the most accurate in describing the diffusion of information in social networks.

Keywords: social networks; mathematical modeling; diffusion of information in networks; cellular automaton; model of adaptive-imitative behavior.

Citation: Gorkovenko D.K. Comparison of epidemic models and cellular automata in

modeling of diffusion of information in social. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems. 2017, Vol. 10, No. 3, Pp. 103–113. DOI: 10.18721/JCSTCS.10309

Введение

Изучение социальных сетей является перспективным научным направлением. В экономическом аспекте это интересно маркетологам и бизнес-аналитикам. Анализ социальных сетей позволяет создать более тесный процесс взаимодействия компании с клиентами. Инструменты анализа позволяют оценить индивидуальные и групповые предпочтения клиентов, выявить тренды интересов и в дальнейшем решать важные стратегические задачи фирмы.

Одной из основных задач сетей является распространение информации [1]. Статьи, обзоры, аудио, видео, короткие сообщения («твиты») и другие виды информации являются основой социальных сетей. Разнообразие информации, возможность создания и продвижения контента, вот что привлекает большие аудитории. И такие аудитории создают клиентскую базу для распространения информации с различными целями: реклама нового продукта, привлечение новых клиентов, удержание старых клиентов и др.

Рассмотрим два основных способа получения информации. Информация может поступить через связи в социальных сетях, либо через средства массовой информации (СМИ). Оценка эффекта от распространения информации через СМИ была сложной задачей. Однако с появлением веб-блогов и других сетевых медиа (новостные порталы, форумы и пр.) анализировать распространение информации стало проще. Вся информация хранится в сети в открытом доступе, пользователи открыто делятся своими мнениями по поводу информации как в текстовом виде, так и через рейтинговые системы (например, отметки «мне нравится», «поделиться»). Это все позволяет более точно изучать процессы диффузии информации, оценивать эффект от распространения. Но моделирование диффузии в социальных сетях остается сложной задачей в связи со сбором большого количества разнообразной информации из разных источников. Кроме того, для исследования необходимо обраба-

тывать и отслеживать элементы социальных сетей, такие как рекомендации, ссылки, теги, сообщения, фразы и «мемы» [2].

Постановка задачи. Цель работы заключается в выборе оптимальной модели, отражающей реальный рост числа «зараженных» пользователей (просмотревших запись) во временном интервале. Такая модель позволит получить информацию об охвате пользователей, скорости распространения информации и каналах распространения.

С помощью данной модели планируется исследовать факторы, влияющие на охват аудитории социальной сети. Например, время публикации, применение виртуального маркетинга на разные сообщества и др.

Исходные данные для исследования. Для сбора данных была выбрана популярная в России социальная сеть «ВКонтакте». Выбрано сообщество с количеством участников 25 775. Открытые данные о связях участников использованы для построения социального графа данного сообщества. Был запущен процесс, который в течение двух суток собирал информацию о новых записях и изменении количества просмотра записей на «стене» сообщества. Основными сущностями для сбора данных выступали: «агент сети», «запись на стене сообщества», «просмотры записи».

Сущность «агент сети» характеризуется данными агента, связями с другими агентами социальной сети и связями с другими сообществами сети.

Сущность «запись на стене сообщества» описана следующими данными: дата создания и текстовая информация.

Сущность «просмотры записи» включает дату обновления и общее количество просмотров.

Полученные данные легли в основу будущего моделирования процессов распространения информации и оценки параметров моделей. Сравнение исходных данных и полученных путем имитационного моделирования показали интересные, с точки зрения исследования, результаты.

Выбор моделей для исследования. В ходе исследования выбраны два типа моделей для моделирования процесса распространения информации: модели эпидемии и модели клеточного автомата.

Модели эпидемии были сформулированы в работе [7] в 1921 г. А в 1965 г. была сформулирована модель Далея-Кендалла для описания распространения слухов. Данный класс моделей до сих пор применяется при моделировании процесса распространения информации (пример – прикладной пакет ISM, система CLAVIRE).

Это объясняется тем, что процесс распространения информации можно сравнить с эпидемией [3, 4]. Особенно в социальных сетях. В связи с отсутствием расстояния между агентами, скорости распространения информации очень высоки (при условии, что информация новая и вызывает интерес), распространение начинается с малых групп и переходит на все большие группы, пока не достигнет пика и не пойдет на спад.

Но модели эпидемии обладают недостатком: они отражают количественное распространение информации и не позволяют получить представление о каналах распространения. Чтобы более точно отразить реальный процесс, была рассмотрена модель клеточного автомата. Клеточный автомат – это дискретная динамическая система, включающая однородные клетки, соединенные друг с другом. Представляет собой сетку произвольной размерности, каждая клетка которой в каждый момент времени может принимать одно из конечного множества состояний, при этом определено правило перехода клеток из одного состояния в другое. Состояние каждой клетки определяется клетками, находящимися в окрестности данной клетки [5]. Такое определение не противоречит определению социальной сети. Исходя из этого можно считать, что модель пригодна для построения социального графа и моделирования социальных процессов.

Оценка параметров моделей. При оценке параметров использован генетический алгоритм. Генетические алгоритмы (ГА) – это область исследований, которая появилась в результате работ Д. Холланда и его коллег.

ГА представляет собой адаптивный поисковый метод, основанный на селекции лучших элементов в популяции. Основой для возникновения и дальнейшего применения таких алгоритмов является модель биологической эволюции и методы случайного поиска [6].

С точки зрения моделирования ГА можно интерпретировать следующим образом: имея начальное множество наборов параметров, путем последовательных преобразований следует получить решение, удовлетворяющее условиям исследования. Само преобразование можно назвать алгоритмом поиска, или генетическим алгоритмом.

ГА обычно состоит из следующих этапов:

- задать целевую функцию (приспособления);
- создание начальной популяции;
- скрещивание;
- мутация;
- вычисление значения целевой функции для всех особей;
- формирование нового поколения;
- если не выполняется условие остановки, то возвращение к третьему этапу.

Функция приспособления – сумма квадратов отклонений между значениями базиса, полученными в результате сбора информации из социальной сети, и значениями, полученными в результате моделирования.

Начальная популяция была задана в виде пяти векторов (x_1, x_1, \dots, x_n) для поиска решения в нескольких направлениях:

1. Значения параметров установлены в крайние значения слева.
2. Значения параметров установлены в крайние значения справа.
3. Значения параметров установлены исследователем.
4. Среднее между вариантом 1 и 3.
5. Среднее между вариантом 2 и 3.

На этапе скрещивания получаем новую популяцию следующим образом: выбираем попарно родителей и вычисляем среднее значение.

На этапе мутации к новой популяции добавляется случайное отклонение.

Программные средства. Для проведения расчетов был выбран язык программиро-

вания Python 3.5. Выбор языка обусловлен его ясным и понятным синтаксисом. Это позволяет сосредоточиться на решении конкретной задачи, а не отвлекаться на особенности языка. Для работы с массивами данных и математическими вычислениями хорошо подходит библиотека NumPy. Для визуализации данных используется библиотека matplotlib, обладающая похожим на Matlab набором функций. Полученные программные средства сформированы в виде набора консольных команд. На вход подается файл с данными социальной сети, тип исследуемой модели. На выходе получаем параметры модели, массив смоделированных данных и визуализацию в виде графика.

Модели эпидемии

Модель SIR. Детерминированная модель эпидемии SIR (susceptible – infected-removed) описывает способ передачи эпидемии от одного индивида (агента) к другому. Процесс имеет параметр затухания [7]. Состояние агента можно описать тремя типами: уязвимое, зараженное, невосприимчивое.

Количество агентов в сети можно выразить как $N = S(t) + I(t) + R(t)$, где $S(t)$ – количество уязвимых агентов, $I(t)$ – количество зараженных агентов, $R(t)$ – количество невосприимчивых агентов. Невосприимчи-

вое состояние можно интерпретировать как потерю интереса к новости и дальнейшее нежелание распространять её. В модели используются следующие параметры: β – средняя частота заражения; γ – постоянная средняя скорость «выздоровления» в единицу времени.

Модель можно представить в виде системы уравнений:

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta SI}{N}, \\ \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I, \\ \frac{dR}{dt} = \gamma I. \end{cases}$$

Используя эту систему уравнений и исходные данные, построим процесс подбора параметров. График полученной имитационной модели отражен на рис. 1. Кривая «зараженных» пользователей очень сильно расходится с данными, полученными из социальной сети. Из этого следует, что данная модель не отражает реальные данные. Поэтому с помощью данной модели невозможно оценить реальный охват пользователей.

Расширенная модель SIR. Социальная сеть обладает изменчивостью во времени. Это означает, что агенты могут присоединиться к сети или покинуть сеть. Обозна-

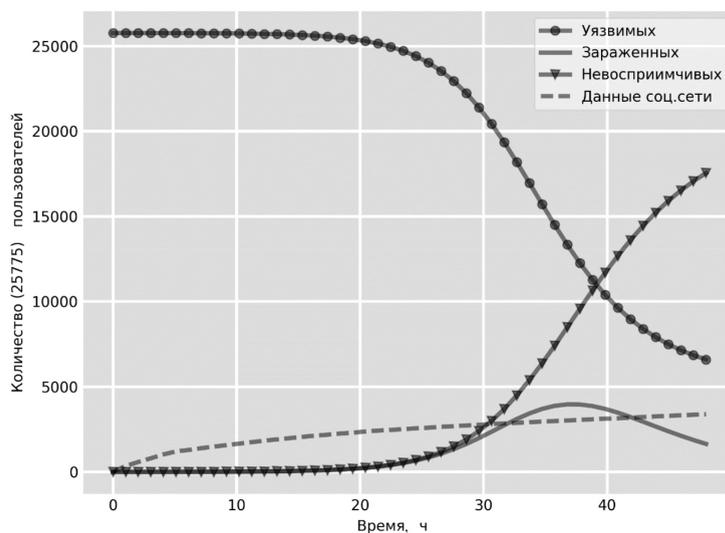


Рис. 1. Модель SIR

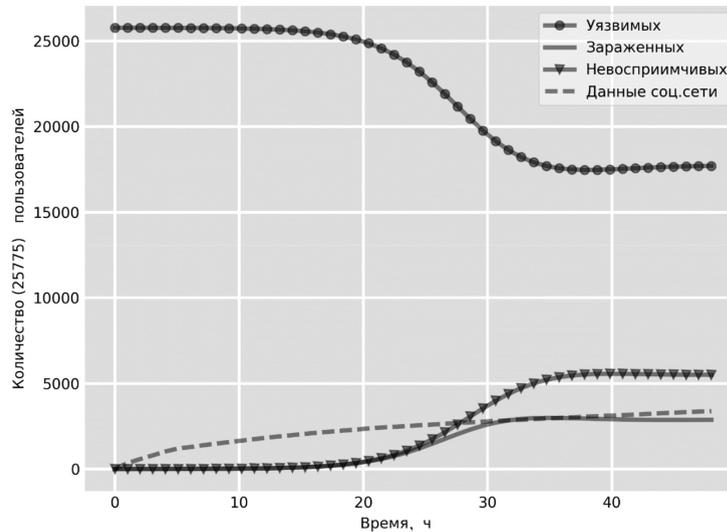


Рис. 2. Расширенная модель SIR

чим параметром μ среднюю частоту присоединения к сети в ед. времени. Параметром δ будем считать среднюю частоту выхода агента из сети в ед. времени. Система уравнений примет следующий вид:

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta SI}{N} + \mu(N - S), \\ \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \delta I, \\ \frac{dR}{dt} = \gamma I - \delta R. \end{cases}$$

Введем параметр α — вероятность перехода из невосприимчивого состояния в уязвимое. Добавим данное условие в систему уравнений модели:

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta SI}{N} + \mu(N - S) + \alpha R, \\ \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \delta I, \\ \frac{dR}{dt} = \gamma I - \delta R - \alpha R. \end{cases}$$

Из графика на рис. 2 видно, что с помощью данной модели можно оценить охват аудитории, но рост числа пользователей во времени она отразить не в состоянии. Большое количество параметров модели дало возможность более гибко построить

кривую роста числа «зараженных» агентов сети.

Модель Далея-Кендалла. Известен метод Далея-Кендалла, описанный в 1965 г. [8], — математическая модель имитации процесса распространения информации (слухов), также называемая ДК моделью. Данная модель делит население на три разные группы:

- группа, которая начинает распространение слуха (U);
- группа, которая после получения слуха продолжает распространять его (V);
- группа, которая после получения слуха принимает решение не распространять его (W).

Модель представлена на рис. 3. N — число участников процесса распространения. Слух распространяется с вероятностью β/N . Степень принятия слуха определена параметром μ . Когда распространитель слухов сталкивается с аудиторией W , то распространение прекращается, и вероятность того, что это произойдет, равна $\frac{\gamma V(V + W)}{N}$.

Слух теряет свою ценность с течением времени. Такая вероятность определяется фактором γ . Это объясняется тем, что слух перестает быть новинкой, или не остается частей, которые можно передать. Модель можно представить в виде уравнений:

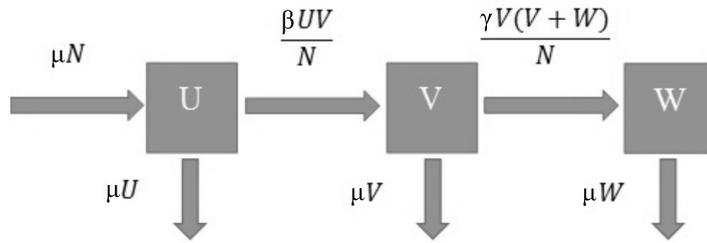


Рис. 3. ДК модель распространения слухов

$$\begin{cases} \frac{dU}{dt} = \mu N - \frac{\beta UV}{N} - \mu U, \\ \frac{dV}{dt} = \frac{\beta UV}{N} - \frac{\gamma V(V+W)}{N} - \mu V, \\ \frac{dW}{dt} = \frac{\gamma V(V+W)}{N} - \mu W. \end{cases}$$

Решение системы можно представить в виде [9]:

$$\begin{cases} U = \frac{\mu N^2}{(\beta V + \mu N)}, \\ V = \frac{N\beta\mu V}{\beta V\gamma + \mu\beta V + \mu^2 N}, \\ W = \frac{\gamma\beta^2 N V^2}{(\beta V + \mu N)(\beta V\gamma + \mu\beta V + \mu^2 N)}. \end{cases}$$

Определим $C_1 = \beta V + \mu N$ и $D_1 = \beta V\gamma$, запишем решение системы в виде:

$$\begin{cases} U = \frac{\mu N^2}{C_1}, \\ V = \left[\frac{\mu N}{D_1 + \mu C_1} \right] \beta V, \\ W = \left[\frac{\gamma N}{C_1(D_1 + \mu C_1)} \right] \beta^2 V^2. \end{cases}$$

Данные моделирования представлены на рис. 4. Хорошо видно, что результаты моделирования схожи с предыдущей моделью. Но в данном случае модель показывает более резкий рост числа «зараженных» в первые сутки. За счет этого отклонение от исходных данных получается выше, чем в

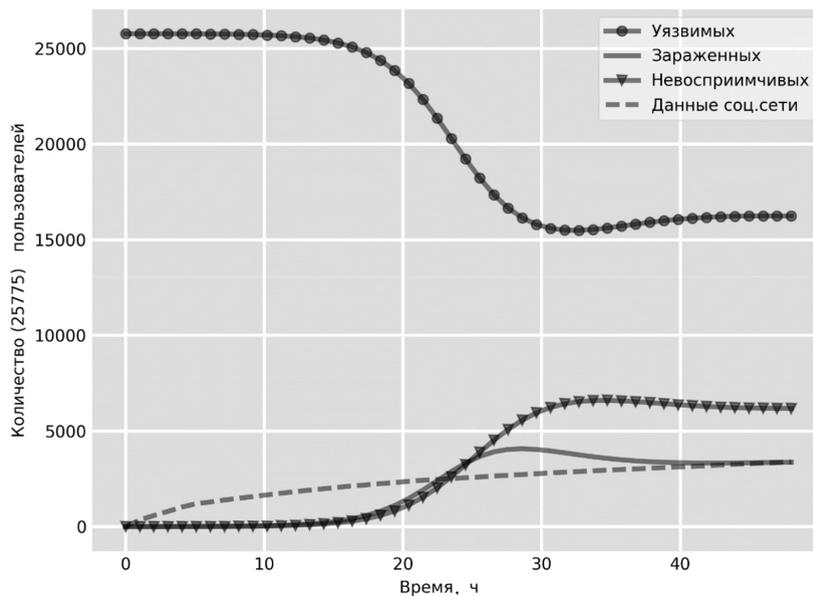


Рис. 4. Модель Далея-Кендалла

расширенной модели SIR.

Модели клеточного автомата

Простейшая модель клеточного автомата. Клеточный автомат – это дискретная динамическая система, включающая однородные клетки, соединенные друг с другом. Все клетки образуют клеточный автомат [5]. Состояние каждой клетки определяется клетками, находящимися в окрестности данной клетки. Набор «ближайших соседей» называется *окрестностью конечного автомата с номером j* . Состояние клеточного автомата j в момент времени $t + 1$ определяется следующим образом:

$$y_j(t + 1) = F(y_j(t), O(j), T),$$

где F – правило, которое может быть выражено (например, в язык булевой алгебры), $O(j)$ – соседи, t – шаг.

Клеточный автомат определен следующими правилами:

- изменение значений каждой клетки происходит одновременно (шагом является изменение единицы времени);
- сеть клеточного автомата является однородной, т. е. правила изменения состояния одинаковы для всех ячеек;
- клетка может влиять только на клетки соседей;

- число состояний клетки конечно.

Теория клеточных автоматов используется для анализа диффузии инноваций. Этот процесс очень похож на распространение новостей в Интернете. Простейшая функция преобразования модели отвечает следующим правилам: индивидuum соответствует одной клетке, которая может принимать два состояния: 1 – новость принята, 0 – новость не принята. Предполагается, что однажды приняв информацию, состояние остается неизменным. Автомат принимает решение о принятии новости, ориентируясь на мнение ближайших соседей: если среди соседей m поддержали инновацию и p – вероятность принятия новости (генерируется в ходе работы модели), тогда если $pm > R$, где R – фиксированное пороговое значение, клетка принимает инновацию. Кроме того, могут быть наложены дополнительные условия на тип новости: клетка располагает свежими новостями (черный цвет), у клетки находится устаревшая информация (серый цвет), клетка не располагает информацией или забыла о ней (белый цвет).

Правила распространения новости:

- 1) сначала каждая клетка закрашена белым цветом, кроме одной черной клетки (которая получила новость);
- 2) белая клетка может изменить цвет на

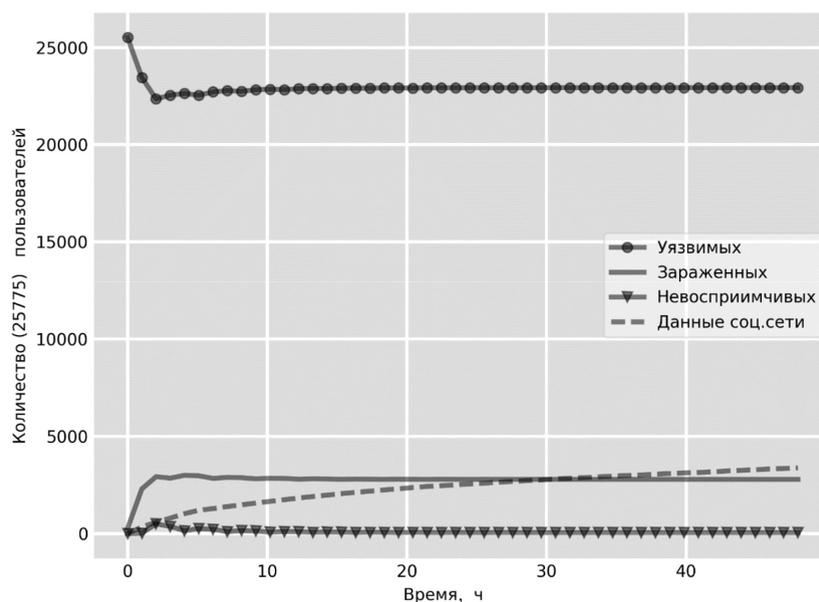


Рис. 5. Модель клеточного автомата

черный или остаться белой (это означает приняла новость или осталась в неведении);

3) белая клетка меняет свой цвет, если условие (1) выполняется в модели распространения диффузий (m – число черных клеток, если $m < 3$, то p увеличивается в 1,5 раза);

4) если ячейка черная и все ячейки вокруг только черные или серые, она меняет свой цвет на серый (новость устаревает);

5) если ячейка серая и ячейки вокруг только черные или серые, то она меняет свой цвет на белый (информация забыта).

На рис. 5 представлены результаты моделирования. Данная модель характеризует резкий рост числа «зараженных» агентов. Полученные данные неточно описывают исходную информацию, полученную из социальной сети.

Модель с порогом. Агент может находиться в активном и неактивном состояниях, причем возможен переход только из неактивного состояния в активное (обратный переход не допускается). Если агент i испытывает влияние a_{ij} каждого своего j -го соседа в сети так, что выполняется условие $\sum_j a_{ij} \leq 1$ (j – активный узел), и становится активным в зависимости от выбранного

им порога $\phi_i \in [0;1]$ (значение может быть фиксированным для всех агентов [10] или может быть выбрано случайным образом в соответствии с некоторым вероятностным распределением [11]), то условие активации $\sum_j a_{ij} \geq \phi_i$ (j – активный узел) [12].

Модель, представленная на рис. 6, смогла отразить рост «зараженных» пользователей на начальном этапе. После роста на начальном этапе (в первые сутки) данная модель приблизилась к пороговому значению и прекратила рост. Итоговый охват оказался далек от первоначальных данных.

Модель независимых каскадов. Данная модель принадлежит к моделям систем взаимодействующих частиц. Состояние агента определяется аналогично узлу в модели с порогом. Если агент i становится активным, то на следующем шаге (строго на следующем) он может активировать соседей с вероятностью p_{ij} [13].

Рисунок 7 отражает результаты моделирования. Можно сказать, что реальный охват был отражен более точно, по сравнению с предыдущей моделью. В моделировании роста на начальном этапе возникли отклонения с базисом.

Модель адаптивно-подражательного поведения (МАПП). МАПП рассмотрена в ра-

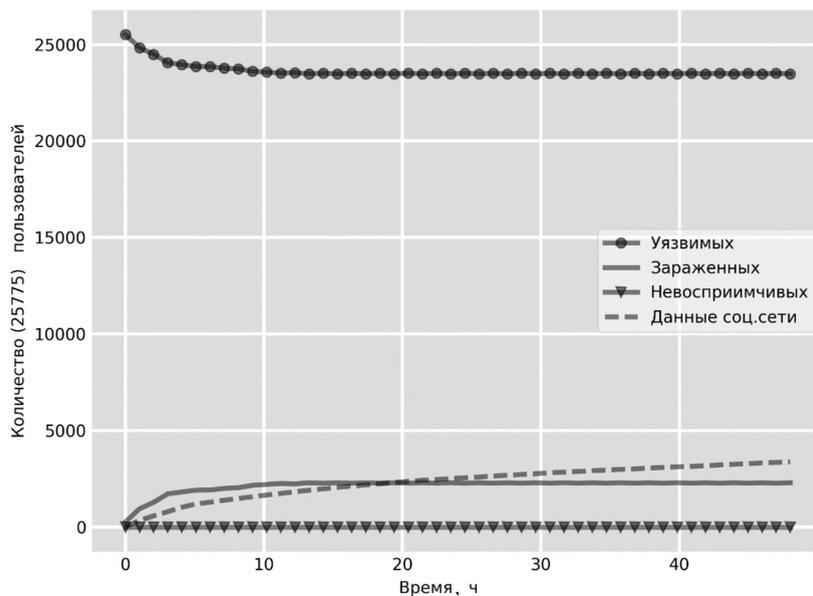


Рис. 6. Модель с порогом

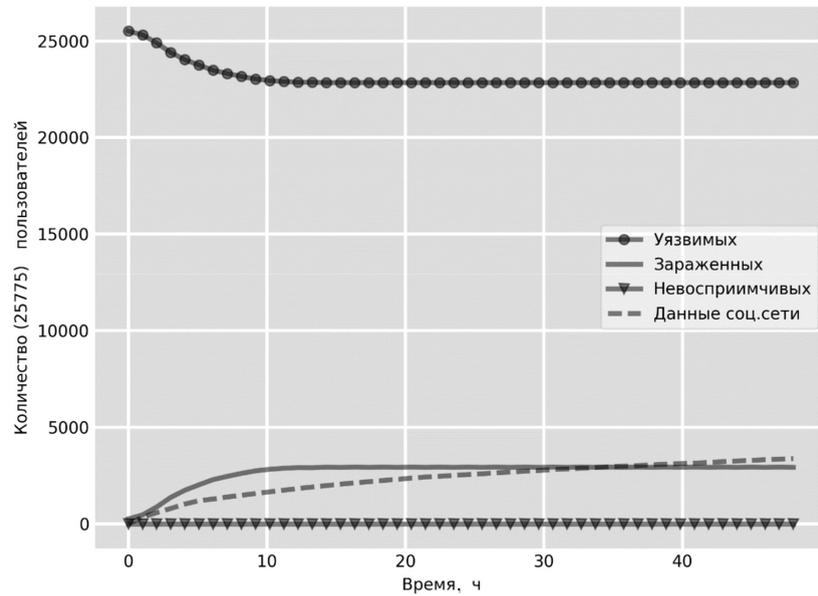


Рис. 7. Модель независимых каскадов

боте [13] и описана в рамках теории игр как $G = \langle J, f_j(\pi), j \in J, \pi \in \Pi \rangle$, где J – множество стратегий участников игры, $\pi = (\pi_j)_{j \in J}$ – распределение игроков по стратегиям, $\Pi = \left\{ \pi \mid \pi_j \geq 0, \sum_{j \in J} \pi_j = 1 \right\}$ – стандартный симплекс, $f_j(\pi)$ – выигрыш игроков, использующих стратегию j . На каждом шаге

агент с индексом i с некоторой интенсивностью переходит в адаптивное состояние, при котором он пересматривает свое мнение (стратегию). В адаптивном состоянии агент i меняет свое мнение на мнение агента k в соответствии с вероятностью $q_{ik} = q_{ik}(f(\pi), \pi)$. Далее сравниваются альтернативная и текущая стратегии. Если выбранная для сравнения стратегия лучше

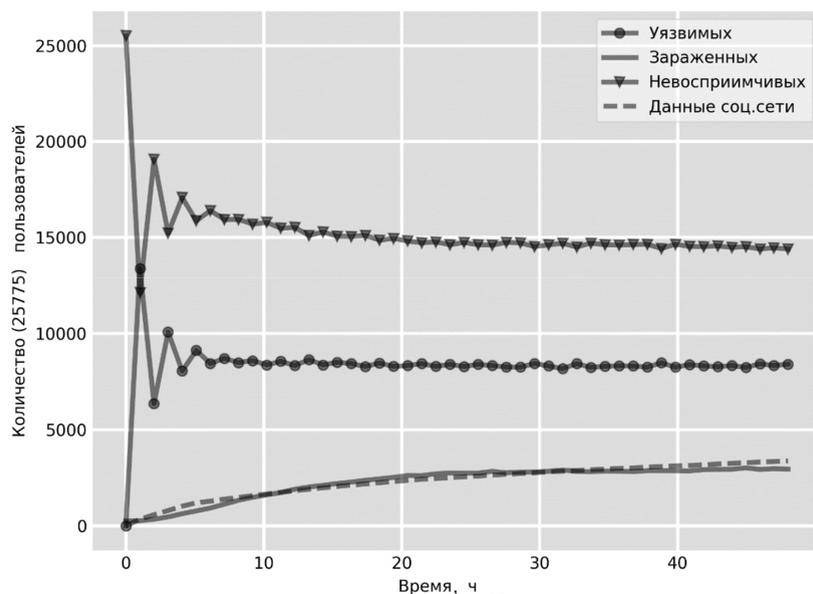


Рис. 8. Модель адаптивно-подражательного поведения (МАПП)

исходной (дает агенту большой выигрыш), то с вероятностью $\gamma_{ik} = \gamma_{ik}(f(\pi), \pi)$ игрок меняет свое мнение.

В ходе моделирования были получены данные, очень близкие к реальным, что детально отражено на графике (рис. 8).

Заключение

Исходя из результатов исследования можно сделать следующие выводы по каждому из типов моделей.

Модели SIR позволяют получить представление об охвате и количественном распределении информации (сколько всего агентов получило информацию), но не позволяют получить представление о каналах распространения информации.

Неприспособленность моделей SIR можно объяснить параметром «выздоровления». При сборе данных нет возможности исправить потерю актуальности новостью. Но такие модели подходят для предварительного подсчета охвата агентов сети. Это позволяет оценить перспективность распространения информации на первоначальном этапе.

Модели клеточного автомата сложны в реализации, их недостаток — длительное время вычисления значений каждого шага. В простейшем приближении не пригод-

ны для моделирования, однако позволяют оценить каналы распространения информации.

Модели клеточного автомата планируется использовать для отражения роста охвата агентов и анализа каналов распространения информации. В результате исследования сделаны выводы о том, что модель адаптивно-подражательного поведения достаточно точно описывает исследуемый процесс. МАПП позволяет описать поведение каждого агента сети, учитывая социальный граф, а также связи между агентами. В данной модели присутствует возможность уточнения стратегий для каждого агента в отдельности, исходя из предпочтений и возможности влияния на агента.

Дальнейшим направлением исследования являются оценка информационного влияния в социальных сетях, изучение каналов распространения информации, возможность влияния на данные каналы, а также рост охвата за счет информационного влияния. Апробированные модели будут сформированы в виде программного пакета для анализа распространения информации в социальных сетях и оценки информационного влияния. Такое программное обеспечение является интересным с точки зрения маркетинга и бизнес-аналитики.

СПИСОК ЛИТЕРАТУРЫ

1. Cha M., Haddadi H., Benevenuto F., Gummadi K.P. Measuring User Influence in Twitter: The Million Follower Fallacy // ICWSM '10. 2010.
2. Leskovec J., Backstrom L., Kleinberg J. Meme-tracking and the dynamics of the news cycle // KDD '09. 2009.
3. Hethcote H.W. The mathematics of infectious diseases // SIAM Review. 2000. No. 42(4). Pp. 599–653.
4. Башабшех М.М., Масленников Б.И., Скворцов А.В. Комбинированная имитационная модель пространственного распространения эпидемических заболеваний по холере на основе вероятностного клеточного автомата // Наукоедение. 2013. № 3 (16).
5. John Von Neumann J., Burks A.W. Theory of Self-Reproducing Automata. University of Illinois Press, Urbana and London, 1966.
6. Емельянов В.В., Курейчик В.В., Курейчик В.М. Теория и практика эволюционного моделирования. М.: Физматлит, 2003. 432 с.
7. Kermack W.O., McKendrick A.G. A Contribution to the Mathematical Theory of Epidemics // Proc. of the Royal Society A: Mathematical, Physical and Engineering Sciences. 1927. No. 115 (772). 700 p. DOI:10.1098/rspa.1927.0118. JSTOR 94815.
8. Daley D.J., Kendall D.G. Stochastic rumors // J. Inst. Math. Appl. 1965. Vol. 142. Pp. 42–55.
9. Isea R., Mayo-Garcia R. Mathematical analysis of the spreading of a rumor among different subgroups of spreaders // Pure and Applied Mathematics Letters. 2015. Pp. 50–54.
10. Friedkin N.E. Structural Cohesion and Equivalence Explanations of Social Homogeneity // Sociological Methods and Research. 1984. No. 12. Pp. 235–261.
11. Васин А.А., Красношеков П.С., Морозов В.В. Исследование операций. М.: Академия, 2008.
12. Робертс Ф.С. Дискретные математические модели с приложениями к социальным,

биологическим и экологическим задачам. М.: Наука, 1986.

13. **Губанов Д.А.** Обзор онлайн-систем

репутации/доверия // Интернет-конференция по проблемам управления. М.: ИПУ РАН, 2009. 25 с.

Статья поступила в редакцию 19.06.2017

REFERENCES

1. **Cha M., Haddadi H., Benevenuto F., Gummadi K.P.** Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM '10*, 2010.

2. **Leskovec J., Backstrom L., Kleinberg J.** Meme-tracking and the dynamics of the news cycle. *KDD '09*, 2009.

3. **Hethcote H.W.** The mathematics of infectious diseases. *SIAM Review*, 2000, No. 42(4), Pp. 599–653, 2000.

4. **Bashabshekh M.M., Maslennikov B.I., Skvortsov A.V.** Kombinirovannaya imitatsionnaya model prostranstvennogo rasprostraneniya epidemicheskikh zabollevaniy po kholere na osnove veroyatnostnogo kletochnogo avtomata [Combined simulation model spatial distribution of epidemic diseases cholera based on probabilistic cellular automata]. *Naukovedeniye*, 2013, No. 3 (16). (rus)

5. **John Von Neumann J., Burks A.W.** *Theory of Self-Reproducing Automata*. University of Illinois Press, Urbana and London, 1966.

6. **Yemelyanov V.V., Kureychik V.V., Kureychik V.M.** *Teoriya i praktika evolyutsionnogo modelirovaniya [Theory and practice of evolutionary modeling]*. Moscow: Fizmatlit Publ., 2003, 432 p. (rus)

7. **Kermack W.O., McKendrick A.G.** A Contribution to the Mathematical Theory of

Epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 1927, No. 115 (772), 700 p. DOI:10.1098/rspa.1927.0118. JSTOR 94815

8. **Daley D.J., Kendall D.G.** Stochastic rumors. *J. Inst. Math. Appl.* 1965, Vol. 142, Pp. 42–55.

9. **Isea R., Mayo-Garcia R.** Mathematical analysis of the spreading of a rumor among different subgroups of spreaders. *Pure and Applied Mathematics Letters*, 2015, Pp. 50–54.

10. **Friedkin N.E.** Structural Cohesion and Equivalence Explanations of Social Homogeneity. *Sociological Methods and Research*, 1984, No. 12, Pp. 235–261.

11. **Vasin A.A., Krasnoshchekov P.S., Morozov V.V.** *Issledovaniye operatsiy [Operations research]*. Moscow: Akademiya Publ., 2008. (rus)

12. **Roberts F.S.** *Diskretnyye matematicheskiye modeli s prilozheniyami k sotsialnym, biologicheskim i ekologicheskim zadacham [Discrete mathematical models with applications to social, biological and ecological problems]*. Moscow: Nauka Publ., 1986. (rus)

13. **Gubanov D.A.** Obzor onlaynovykh sistem reputatsii/doveriya [Review of online reputation / trust systems]. *Internet konferentsiya po problemam upravleniya [Internet Conference on Governance issues]*. Moscow: IPU RAN, 2009, 25 p. (rus)

Received 19.06.2017

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

ГОРКОВЕНКО Дмитрий Константинович

GORKOVENKO Dmitry K.

E-mail: gorkovenko.dmitry@gmail.com

НАУЧНОЕ ИЗДАНИЕ
«НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ
САНКТ-ПЕТЕРБУРГСКОГО
ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА.
ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ»
«ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL.
COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS»

Том 10, № 3, 2017

Учредитель – Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Санкт-Петербургский государственный политехнический университет»

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор).
Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

Редакция журнала

д-р техн. наук, профессор *А.С. Коротков* – главный редактор
Е.А. Калинина – литературный редактор, корректор
Г.А. Пышкина – ответственный секретарь, выпускающий редактор

Телефон редакции (812)552-62-16, 297-18-21

E-mail: infocom@spbstu.ru

Компьютерная верстка *А.Н. Смирнов*

Перевод на английский язык *А.С. Колгатина*

Лицензия ЛР № 020593 от 07.08.97

Подписано в печать 29.09.2017. Формат 60×84 1/8. Бум. тип. № 1.
Печать офсетная. Усл. печ. л. 13,48. Уч.-изд. л. 13,48. Тираж 1000. Заказ

Санкт-Петербургский политехнический университет Петра Великого
Издательство Политехнического университета
член Издательско-полиграфической ассоциации университетов России
Адрес университета и издательства: 195251, Санкт-Петербург, ул. Политехническая, д. 29.

УСЛОВИЯ ПУБЛИКАЦИИ СТАТЕЙ

в журнале «Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление»

1. ОБЩИЕ ПОЛОЖЕНИЯ

Журнал «Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Телекоммуникации. Управление» является периодическим печатным научным рецензируемым изданием. Зарегистрировано Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19 октября 2012 г. С 2008 года выпускается в составе сериального периодического издания «Научно-технические ведомости СПбГПУ» (ISSN 1994-2354).

Издание с 2002 года входит в Перечень ведущих научных рецензируемых журналов и изданий (перечень ВАК) и принимает для печати материалы научных исследований, а также статьи для опубликования основных результатов диссертаций на соискание ученой степени доктора наук и кандидата наук по следующим основным научным направлениям: **ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА, РАДИОТЕХНИКА И СВЯЗЬ, ЭЛЕКТРОНИКА, ИЗМЕРИТЕЛЬНАЯ ТЕХНИКА, УПРАВЛЕНИЕ В СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ СИСТЕМАХ**. Научные направления журнала учитываются ВАК Минобрнауки РФ при защите докторских и кандидатских диссертаций в соответствии с Номенклатурой специальностей научных работников.

Сведения о публикациях представлены в РИНЦ, в Реферативном журнале ВИНИТИ РАН, в международной справочной системе «Ulrich's Periodical Directory».

Периодичность выхода журнала – 4 номера в год.

2. ТРЕБОВАНИЯ К ПРЕДОСТАВЛЯЕМЫМ МАТЕРИАЛАМ

2.1. Оформление материалов

1. Рекомендуемый объем статей для авторов с ученой степенью доктора наук, званием профессора, соискателей ученой степени доктора наук (докторантов) 12–20 страниц формата А-4 с учетом графических вложений. Количество графических вложений (диаграмм, графиков, рисунков, таблиц, фотографий и т. п.) не должно превышать 4.

2. Рекомендуемый объем статей для преподавателей, авторов без ученой степени, соискателей ученой степени кандидата наук – 8–15 страниц формата А-4; аспирантов – 8 страниц формата А-4 с учетом графических вложений. Количество графических вложений (диаграмм, графиков, рисунков, таблиц, фотографий и т. п.) не должно превышать 3.

3. Авторы должны придерживаться следующей обобщенной структуры статьи: вводная часть (0,5–1 стр., актуальность, существующие проблемы); основная часть (постановка и описание задачи, изложение и суть основных результатов); заключительная часть (0,5–1 стр., предложения, выводы), список литературы (оформление по ГОСТ 7.05.-2008).

4. Число авторов статьи не должно превышать трех человек.

5. Набор текста осуществляется в редакторе **MS Word**, формул – в редакторе **MathType**. Таблицы набираются в том же формате, что и основной текст.

6. Шрифт – **TNR**, размер шрифта основного текста – 14, интервал – 1,5; таблицы большого размера могут быть набраны 12 кеглем. Параметры страницы: поля слева – 3 см, сверху, снизу – 2,5 см, справа – 2 см, текст размещается без переносов. Абзацный отступ – 1 см.

2.2. Предоставление материалов

Вместе с материалами статьи должны быть обязательно предоставлены:

- номер УДК в соответствии с классификатором (в заголовке статьи);
- аннотация на русском и английском языках;
- ключевые слова (5–7) на русском и английском языках;
- сведения об авторах на русском и английском языках: ФИО, место работы, должность, ученое звание, ученая степень, контактные телефоны, e-mail;
- аспиранты представляют документ отдела аспирантуры, заверенный печатью;
- акт экспертизы о возможности опубликования материалов в открытой печати.

С авторами статей заключается издательский лицензионный договор.

Предоставление всех материалов осуществляется в электронном виде через личный кабинет **ЭЛЕКТРОННОЙ РЕДАКЦИИ** по адресу <http://journals.spbstu.ru>

2.3. Рассмотрение материалов

Предоставленные материалы (п. 2.2) первоначально рассматриваются редакционной коллегией и передаются для рецензирования. После одобрения материалов, согласования различных вопросов с автором (при необходимости) редакционная коллегия сообщает автору решение об опубликовании статьи. В случае отказа в публикации статьи редакция направляет автору мотивированный отказ.

При отклонении материалов из-за нарушения сроков подачи, требований по оформлению или как не отвечающих тематике журнала материалы не публикуются и не возвращаются.

Редакционная коллегия не вступает в дискуссию с авторами отклоненных материалов.

Публикация материалов аспирантов очной бюджетной формы обучения осуществляется бесплатно в соответствии с очередностью.

При поступлении в редакцию значительного количества статей их прием в очередной номер может закончиться **ДОСРОЧНО**.

Более подробную информацию можно получить:

на сайте журнала <http://ntv.spbstu.ru>

по телефону редакции +7(812) 552-62-16 с 10⁰⁰ до 18⁰⁰ Галина Александровна

или по e-mail: infocom@spbstu.ru