МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ»

Вестник СибГУТИ Том 18, № 3, 2024

Выпускается ежеквартально, выходит с 2007 г.

Учредитель –
федеральное государственное бюджетное образовательное учреждение высшего образования
«Сибирский государственный университет телекоммуникаций и информатики»

Председатель редакционного совета А. Н. Фионов, д.т.н., проф.

Редакционный совет:

С. С. Абрамов, д.т.н., доц.
В. Б. Барахнин, д.т.н., доц.
В. М. Белов, д.т.н., проф.
В. Н. Васюков, д.т.н., проф.
В. Ю. Васильев, д.х.н., доц.
Н. И. Горлов, д.т.н., проф.
Н. Л. Казначеева, д.э.н., доц.
В. С. Канев, д.т.н., проф.
А. И. Карпович, д.э.н., проф.
М. Г. Курносов, д.т.н., проф.
В. В. Лебедянцев, д.т.н., проф.
А. В. Лихачёв, д.т.н.
С. Н. Мамойленко, д.т.н., доц.
О. Г. Мелентьев, д.т.н., проф.
Р. В. Мещеряков, д.т.н., проф.
И. Г. Неизвестный, д.ф.-м.н., чл.-корр. РАН
С. Н. Новиков, д.т.н., доц.
В. И. Носов, д.т.н., проф.
С. В. Поршнев, д.т.н., проф.
А. С. Родионов, д.т.н., доц.
А. И. Романенко, д.ф.-м.н., проф.
Б. Я. Рябко, д.т.н., проф.
И. И. Рябцев, д.ф.-м.н., чл.-корр. РАН
Э. Сименс, д.т.н.
О. В. Стукач, д.т.н., доц.
В. К. Трофимов, д.т.н., проф.
А. И. Фалько, д.т.н., проф.
С. В. Федоренко, д.т.н., доц.
Б. Г. Хаиров, д.э.н., доц.
С. В. Шидловский, д.т.н.
Ю. И. Шокин, д.ф.-м.н., акад. РАН
В. П. Шувалов, д.т.н., проф.

Редакция:

А. Н. Фионов (главный редактор), М. Ю. Галкина (заведующая редакцией),
Н. А. Двуреченская (технический и литературный редактор, компьютерная вёрстка),
Т. А. Алфёрова (лингвист-корректор).

Адрес редакции и издателя

630102, г. Новосибирск, ул. Кирова, д. 86 e-mail: vestnik@sibguti.ru

Информация о журнале доступна в сети Internet по адресу https://vestnik.sibsutis.ru.

Журнал включён в Перечень ВАК российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёных степеней доктора и кандидата наук.

Журнал зарегистрирован Федеральной службой по надзору за соблюдением законодательства в сфере массовых коммуникаций и охране культурного наследия, свидетельство о регистрации ПИ № ФС77-25835 от 29.09.2006. ISSN 1998-6920. Подписной индекс в каталоге «Урал-Пресс» – 82519.

Дата выхода в свет 04.10.2024. Отпускная (редакционная) цена 750 руб.
Отпечатано в типографии Альфа по адресу: 630015, г. Новосибирск, ул. Шишкина, д. 3.
Бумага офсетная, формат А4. Тираж 300 экз.

© СибГУТИ, 2024 г.

СОДЕРЖАНИЕ

А. О. Логинова
Описательная модель социального бота, учитывающая его потенциал нарушителя информационно-психологической безопасности …… 3

Е. В. Вершинина, О. М. Зверева, С. Н. Шабунин
Методика формирования проектных команд на основе ролевой модели Белбина ……. 14

И. Л. Виноградова, Е. Ю. Головина
Анализ эффективности генерирования и преобразования оптических ФИМД-сигналов на базе классического RoF-сегмента, основанного на применении двух лазеров …… 28

А. М. Кальней, А. С. Родионов
Одна задача структурной оптимизации сетей передачи данных ……. 45

В. В. Шубин
Отказы интегральных схем, вызванные воздействием эффектов электромиграции и антенны …… 57

А. В. Токарева, Д. В. Кручинин
Модификация метода идентификации и прослеживаемости сложных технических изделий с применением алгоритмов комбинаторной генерации на основе деревьев И/ИЛИ …… 73

П. А. Сеченов, И. А. Рыбенко
Сравнение программных реализаций методов решения СЛАУ в задаче нахождения равновесного состава сложной многокомпонентной гетерогенной системы …….. 86

Т. Л. Самков
Финансовый баланс экономики региона …… 99

А. А. Березкин, А. А. Ченский, Р. В. Киричек
Исследование границ интенсивности видеопотока при FPV-управлении БПЛА в режиме предсказания кадров. Часть I: модели и методы ……... 115

DOI: 10.55648/1998-6920-2024-18-3-3-13 УДК 004.5; 004.8:004.9

Описательная модель социального бота, учитывающая его потенциал нарушителя информационно-психологической безопасности *

А. О. Логинова

Московский государственный лингвистический университет (ФГБОУ ВО МГЛУ)

Аннотация: Цель настоящего исследования заключается в создании описательной модели социального бота, учитывающей его потенциал нарушителя информационно-психологической безопасности. Исследование проводилось на примере американского политического дискурса. В основу модели легли результаты комплексного анализа неразмеченного корпуса текстов постов социальных ботов в Twitter. Исследуемый корпус текстов представляет собой совокупность текстов постов на английском языке аккаунтов в социальной сети, которые ранее были определены исследователями как социальные бот-аккаунты, задействованные в предвыборной кампании кандидатов в президенты. Исследование мотивировано отсутствием исчерпывающего определения понятия «социальный бот». Разработанная модель бота формирует новый подход к раскрытию сути данного понятия. В настоящей статье социальный бот рассматривается как нарушитель информационно-психологической безопасности пользователя интернет-средств массовой коммуникации.

Ключевые слова: социальный бот, модель нарушителя, информационно-психологическая безопасность, интернет-средства массовой коммуникации, социальные сети.

Для цитирования: Логинова А. О. Описательная модель социального бота, учитывающая его потенциал нарушителя информационно-психологической безопасности // Вестник СибГУТИ. 2024. Т. 18, № 3. С. 3–13. https://doi.org/10.55648/1998-6920-2024-18-3-3-13.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Логинова А. О., 2024

Статья поступила в редакцию 22.01.2024;
переработанный вариант – 25.02.2024;
принята к публикации 27.02.2024.

1. Введение

Сегодня, когда доступ к различного рода медиаконтенту есть у всех интернет-пользователей, а их число ежегодно растет [1], практика информационного воздействия обретает совершенно иные масштабы. Методы информационного воздействия в равной мере используются как для достижения целей маркетинга, так и для ведения боевых действий, которые, как отмечают эксперты, всё чаще ведутся в киберпространстве.

Одним из современных инструментов информационного воздействия, бесспорно, является социальный бот (далее – бот, социальный бот, бот-аккаунт, автоматизированный социальный актор) [2, 3]. В литературе существует большое разнообразие определений понятия со-

* Статья подготовлена при поддержке Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации (научный проект № 14/23-К).

циального бота, но они дают лишь частичное понимание того, что на самом деле скрывается за ним.

Цель настоящего исследования заключается в создании модели социального бота на примере американского политического дискурса.

В работе применяется метод описательного моделирования социального бота как нарушителя информационно-психологической безопасности пользователей интернет-средств массовой коммуникации (социальные сети, блоги, форумы и др.) (далее – интернет-СМК).

Материалом исследования послужил неразмеченный корпус текстов коротких электронных сообщений пользователей социальной сети Twitter на английском языке, собранный из веб-архива, которые были заблокированы модераторами Twitter как аккаунты социальных ботов, задействованные в предвыборной кампании кандидатов в президенты США в 2015–2016 годах.

Такой выбор материала для исследования был обусловлен рядом аргументов:

– в ходе избирательной кампании в качестве основного канала связи с электоратом впервые была использована социальная сеть [4, 5], поэтому в архиве Интернета [6] на сегодняшний день хранится достаточно большой объём текстовой информации, в частности, содержание постов главных кандидатов в президенты, которые могут быть использованы для анализа;

– доказано, что при реализации политической пропаганды использовались социальные боты [4, 7];

– коллективы исследователей, занимающиеся вопросом обнаружения социальных ботов, сформировали определения понятия «социальный бот», отличающиеся в части описания природы происхождения автоматизированного социального актора [8 – 10]. Этот факт поспособствовал формированию гипотезы о том, что социальные боты имеют не только различную природу происхождения, но и разный потенциал нарушителя информационно-психологической безопасности. До сих пор данное предположение не подвергалось проверке.

2. Используемые определения понятия «социальный бот»

Действующие нормативные правовые акты не дают определения понятия «социальный бот», вместе с тем различные исследовательские коллективы приводили свои определения [8 – 10]. В совокупности они формируют следующее представление о социальном боте в целом: это в первую очередь аккаунт в социальной сети, который управляется с помощью специального программного обеспечения и контролируется человеком. Результатом работы такого аккаунта является создание и распространение заданного контента среди участников некоторого интернет-сообщества в социальной сети.

Действия таких ботов подвергают угрозам информационно-психологическую безопасность (далее – ИПБ) реальных пользователей интернет-СМК. Отсутствие возможности у пользователя достоверно идентифицировать коммуниканта – причина реализации угроз ИПБ [11].

Результаты проведённого нами исследования текстов опубликованных сообщений, постов, ботов в социальной сети Twitter идут вразрез с содержанием устоявшегося определения в части, касающейся способа управления бот-аккаунтом и генерации текстов.

3. Систематизация наблюдений

Значения относительной частоты употребления конкретных единиц языка в исследуемых корпусах текста, авторства людей и социальных ботов, полученные в результате предшествующих этапов исследования на материале американского политического дискурса, относящегося к президентской гонке 2015 – 2016 годов в США [12], демонстрируют незначительные отличия текстовых сообщений социальных ботов (в сложившемся понимании) от сообщений людей.

При этом рассмотрение сообщений ботов в индивидуальном порядке позволяет выделить некоторые синтаксические особенности постов, опубликованных несколькими аккаунтами, определённых исследователями как социальные боты. Выявленные особенности обнаруживают несоответствие действительности устоявшихся определений понятия «социальный бот».

Так, например, при создании текстов постов, отражённых в табл. 1, использовались приёмы суггестии [13]. Данные приёмы не могли быть реализованы только лишь посредством программного обеспечения, генерирующего текстовые сообщения для социального бота, включая возможности ChatGPT, без редакторской правки человеком. Это может свидетельствовать о том, что тексты, публикуемые вышеуказанными аккаунтами, составлены человеком или группой людей, имеющими обширные знания в области социальной психологии и социальной инженерии. Отметим также, что чат-бот ChatGPT был анонсирован только в 2020 году, т.е. гораздо позже президентской гонки между Х. Клинтон и Д. Трампом.

Некоторая доля постов бот-аккаунтов, например, @Marycar08639249, @PatDollard и @u_edilberto, состоит из репостов текстов других аккаунтов, принадлежащих действительным пользователями социальной сети или также являющихся ботами.

Таблица 1. Примеры использования приёмов суггестии в сообщениях социальных ботов

№ п/п	Пример поста социального бота (сохранены авторские орфография и пунктуация)	Перевод выделенных фрагментов	Приём суггестии
1.	«*@63df8aabe784aebc5674b36a34c14dae77bfaa89e1d9bafa5a485ed32f6e2834* If Hillary becomes president I'm moving to Benghazi, that way I'll know she will leave me alone! **#Hillary4Prison** #TrumpForPresident»	#ХиллариВтюрьму	навешивание ярлыков: #Hillary4Prison
2.	«*@Marycar08639249* RT @AlwaysActions: Powerful response to obama's "insult" speech by Donald Trump supporting @USArmy @AdBell45 **#VoteTrump2016 #Trump2016** htt ...» [7]	#ГолосуйЗаТрампа2016 #Трамп2016	повтор информации
3.	«*@natespuewell* #NeverTrump **Those fake, nonsense polls are actually real, good polls, Trump's spokesman insists — Campaign of lies** `https://t.co/Mvja0PPeaH`» [7]	«Эти фальшивые, бессмысленные результаты голосований на самом деле настоящие, правильные результаты, настаивает пресс-секретарь Трампа – Кампания лжи»	утверждение
4.	«*@jeannemccarthy0 RT* @realDonaldTrump: **The best thing to die in 2016?** Political correctness.We have been set free from the intolerant left. Let's thank `https://t.co/cO9gm7aAcZ` » [7]	«Лучшее, за что можно умереть в 2016?»	постановка риторических вопросов
5.	*@pavegecko01* #**Hillary can't walk down the stairs by herself** `https://t.co/8DOZSwaHNm`	«Хиллари не может спуститься по лестнице самостоятельно»	создание послеобраза

Информация о репосте сообщается подписчикам любого аккаунта в Twitter посредством следующего набора символов: «RT @nickname:», где RT сообщает о том, что имеет место репост, а @nickname – указание наименования аккаунта, с которого был сделан репост, после символа «:» записан текст оригинального поста, например:

– @Marycar08639249: RT @AlwaysActions: Powerful response to Obama's "insult" speech […];
– @PatDollard: RT @DesertRiver: Oh Goody... […];
– @u_edilberto: RT @WeNeedHillary: Polls Are All Over t... […].

Однотипные операции репоста могли выполняться человеком, выполняющим простые повторяющиеся операции, или путём автоматизации работы такого аккаунта посредством программного обеспечения.

Некоторые сообщения исследуемого корпуса, опубликованные разными пользователями, идентифицированные исследователями как боты, в точности повторяют друг друга без указания информации о репосте, как если бы имело место легитимное копирование поста другого автора. Например:

– @CloudBeDelusion: Fox’s Shepard Smith Call’s Bulls*t […];
– @ProletariatStriving: Fox’s Shepard Smith Call’s Bulls*t […];
– @FillingDCSwamp: Pattern Of Hacking Preceded Attendee Of Donald Trump’s Camp […];
– @ProletariatStriving: Pattern Of Hacking Preceded Attendee Of Donald Trump’s Camp […].

Такая особенность текстов постов может свидетельствовать о том, что данные аккаунты входят в состав бот-нета, владелец которого может являться как продвинутым специалистом, способным развернуть подобный бот-нет, так и организованной группой людей, в которой каждый участник – узел бот-нета или же усилия каждого участника направлены на то, чтобы развернуть свою сеть ботов в целях достижения общей цели, например, на политической арене.

Полученные результаты позволяют сделать вывод о том, что в некоторых случаях задача обнаружения социального бота может трансформироваться в задачу обнаружения аккаунта человека, действующего как бот. В этом случае обнаружение деятельности социального бота – поиск конкретной языковой личности по особенностям создаваемого текста.

Подчеркнём, что сообщения ChatGPT, способного генерировать ответы на запросы, относящиеся к различным сферам жизни, в режиме диалога сегодня не могут быть приравнены к сообщениям людей, поскольку восприятие текста человеком в целом зависит не только от корректности грамматических конструкций и соответствия ответа заданному вопросу. Текст, созданный человеком, – результат социальной рефлексии.

В настоящее время пользователями ChatGPT обнаружена возможность чат-бота имитировать эмпатию посредством языковых конструкций при генерации текстов. Данный факт был установлен эмпирическим путём. В своих экспериментах различные пользователи ChatGPT формулировали запрос к чат-боту как к психотерапевту с указанием желаемых характеристик тона излагаемых рекомендаций, характеристик предлагаемых действий для решения проблемы, ограничений и других параметров ответов ChatGPT. Пользователи отмечали, что, несмотря на спокойный тон сообщений и проявления внимательности посредством языковых конструкций, сообщения ChatGPT в значительной мере отличались от речи психотерапевта. Сравнению подвергается именно качество диалога, а не уровень включенности в диалог и применимости получаемых рекомендаций. При общем положительном впечатлении от диалога в нём читалась цепочка: проблема – решение – проблема. Данные ботом советы были обезличены и в большинстве случаев лишены конкретики, т.е. не разрешали частной проблемы обратившегося за консультацией человека. В ответах содержалась общая информация, представляющая собой хорошо структурированную компиляцию информации по некоторому заданному вопросу. Именно отсутствие передачи личного отношения к проблеме с использованием средств языка отличало тексты ChatGPT, генерирующего тексты достаточно высокого качества, от текстов, созданных человеком.

Таким образом, предположив существование чат-бота ChatGPT в 2015 году, отметим, что даже сегодняшний уровень развития данной технологии и предоставляемые на сегодняшний день возможности противоречат предположению о том, что тексты, сгенерированные с помощью ChatGPT, могут оказывать суггестивное воздействие на человека. Это связано с тем, что манипуляционное воздействие методом суггестии характеризуется воздействием на чувства человека, которые блокируют его способность критически оценивать ситуацию. В наблюдениях пользователей ChatGPT не отмечалось ощущения полного погружения коммуниканта, в роли которого выступал ChatGPT, использующий речевую модель психотерапевта, в проблему «пациента», вместо этого прослеживалось использование шаблонов в формулировках ответов. Это не позволяло пользователям ChatGPT избавиться от ощущения общения с ботом, что доказывает несостоятельность предположение об использовании методов суггестии при генерации текстов социальных ботов только с помощью ChatGPT без коррекции текста человеком, компетентным в сфере социальной психологии, в частности социальной инженерии.

4. Формирование модели социального бота

Опираясь на данные, полученные в ходе изучения корпуса текстов постов социальных ботов в американском политическом дискурсе, построим модель социального бота по принципу модели нарушителя информационной безопасности (рис. 1).

Модель, составленная на материале текстов ботов в политическом дискурсе, отражает основные общие характеристики бот-аккаунта в социальной сети, а именно:

1. Правомочие пользования аккаунтом в социальной сети: использование украденного аккаунта в социальной сети / использование вновь созданного аккаунта.

2. Способ управления контролируемым аккаунтом: аккаунт управляется с помощью программного обеспечения (далее – ПО) / аккаунт управляется человеком.

Отметим, что аккаунт, управляемый с помощью ПО, не может быть полностью автономным (в модели установленный факт обозначен зависимостью). Это утверждение справедливо по ряду причин. ПО, позволяющее автоматизировать работу аккаунта в социальной сети, не является предустановленным, решение об использовании или об отказе в использовании любого ПО для работы аккаунта принимается его владельцем.

Человек может выполнять функции автомата, реализовывать простые однотипные действия, например, делать репосты или копировать тексты постов в нарушение авторских прав.

В описанном случае человек может являться как владельцем процесса и устанавливать цели, которые реализуются путём массовой публикации контента определённого содержания, так и являться исполнителем. Человек может быть как владельцем бот-аккаунта и управлять им единолично, так и быть частью организованной группы людей.

3. Выполняемые с помощью управляемого аккаунта действия.

В наборах выполняемых действий наблюдается градация: сложность выполняемых действий увеличивается со сложностью организации нарушителя. Так, для аккаунта, управляемого с помощью ПО, и аккаунта, управляемого человеком, но с низким потенциалом нарушителя, набор действий прост и заключается в копировании постов, копировании текстов постов других аккаунтов, генерации и публикации новых текстов постов. При этом для аккаунта, управляемого человеком или группой лиц с высоким потенциалом нарушителя, указано создание бот-нета, генерирующего посты с использованием всех доступных инструментов. Бот-нет при этом может состоять как из аккаунтов, управляемых только с помощью ПО, так и из аккаунтов, управляемых людьми, или комбинации аккаунтов разной природы.

Рис. 1. Описательная модель социального бота

4. Потенциал нарушителя ИПБ: низкий, средний, высокий, – комплексный критерий оценки нарушителя с точки зрения ресурсов, которые он может использовать для реализации цели. В качестве ресурсов, определяющих потенциал нарушителя, в данной модели были приняты: вычислительная мощность, уровень финансового обеспечения, уровень знаний в информационных технологиях, уровень знаний в социальной психологии. Такой выбор обусловлен тем, что:

– количество генерируемых постов в единицу времени, количество поддерживаемых диалогов, масштабируемость бот-нета, т.е. допустимое число управляемых аккаунтов в бот-нете, зависят от производительности контролирующей системы;

– уровень знаний информационных технологий определяет тип действий владельца аккаунта: наверстывающий или опережающий. Высокий уровень знаний позволяет своевременно внедрять доступные инструменты в работу;

– уровень финансового обеспечения является определяющим фактором для увеличения производительности: будь то закупка дополнительного оборудования для увеличения вычислительной мощности оборудования или оплата труда людей, генерирующих и публикующих тексты постов;

– успех применения социальных ботов объясняется эффективностью методов социальной инженерии. Их использование в коммуникации в социальных сетях позволяет увеличить длину пути социального графа – увеличить число действительных пользователей, подвергающихся воздействию. Уровень знаний нарушителей в области социальной психологии, в частности в социальной инженерии, во многом определяет успешность реализуемой атаки.

5. Цели, преследуемые лицом, управляющим бот-аккаунтом, или владельцем бот-аккаунта, имеют прямо пропорциональную зависимость с потенциалом нарушителя. В модели представлены индивидуальные цели, или цели исполнителей (создание положительного имиджа заказчика в интернет-СМК), локальные (увеличение клиентской базы) и глобальные цели (дискредитация оппонентов, увеличение электората).

6. Вид нарушителя для каждого уровня потенциала нарушителя ИПБ.

Набор предположений о располагаемой вычислительной мощности нарушителя информационно-психологической безопасности, уровне финансового обеспечения, его уровне знаний в области информационных технологий и в области социальной психологии, сформированный на основе наблюдений, позволил составить общую характеристику нарушителей, предположить род их деятельности.

В модели представлены примеры нарушителей, которые могут работать как поодиночке, так и в составе организованных групп. Так, примерами лиц, управляющих аккаунтами с использованием или без использования специального ПО с низким потенциалом, являются человек, выполняющий низкоквалифицированную работу, или начинающий SMM-специалист, со средним потенциалом – продвинутый SMM-специалист, с высоким потенциалом – организованные группы лиц, а именно: конкурирующий внутригосударственный политический штаб или зарубежный политический штаб. В ряд нарушителей с высоким потенциалом при необходимости можно включить представителей разведывательных организаций.

На основе приведённой модели социального бота попытаемся сформулировать новое определение понятия «социальный бот»: это аккаунт в интернет-СМК, владельцем которого может являться как физическое лицо, так и организованная группа лиц, который управляется человеком с использованием специального ПО, позволяющего автоматизировать процедуры рассылки, копирования и генерации новых текстов сообщений и (или) постов. Аккаунт может быть автономным или являться частью бот-нета. Цель использования таких аккаунтов – создание и массовое распространение контента, оказывающего информационно-психологическое воздействие на участников сообщества в некотором интернет-СМК.

5. Перспективы исследования

Учитывая сформированную модель, можно определить общий вид нарушителя ИПБ, социального бота, на обнаружение которого направлены существующие исследования. В качестве примеров рассмотрим работы российских исследователей.

В работе Менщикова А. А. [14] используется методика, основанная на построении графа пользовательского поведения: производится анализ действий пользователей и их последовательности в социальных сетях. Разработанная Менщиковым А. А. методика направлена на обнаружение веб-роботов, т.е. аккаунтов со встроенным ПО, позволяющим выполнять заданные действия. Следовательно, рассматриваемая методика подходит для обнаружения аккаунтов, управляемых с помощью специального ПО, имеющих низкий потенциал нарушителя.

Чесноковым В. О. в целях обнаружения интернет-ботов было проведено исследование социальных графов, взаимосвязей пользователей социальных сетей друг с другом с последующим выделением сообществ в социальной сети [15]. В сформированной модели представлен ряд правил, по которым должна производиться фильтрация аккаунтов в социальной сети с целью выделения среди них бот-аккаунтов. С точки зрения представленной модели социального бота методика обнаружения ботов, предложенная Чесноковым В. О., направлена на обнаружение аккаунтов как управляемых с помощью специального ПО, так и управляемых человеком. Нарушители ИПБ при этом имеют низкий и средний потенциалы.

Обнаружение социальных ботов с высоким потенциалом нарушителя, представленных аккаунтами, управляемыми людьми, может рассматриваться как последовательная работа по обнаружению единичных узлов бот-нета и установлению взаимосвязей между ними.

Попытка разработки методики выявления бот-нета предпринята коллективом исследователей в составе Васильковой В. В. и Легостаевой Н. И [16]. Методика, предложенная исследователями, включает комплекс из анализа всплесков публикационной активности пользователей социальной сети, анализа постов, профайлингов бот-аккаунтов, статистического анализа текстовых данных, анализа топологии бот-нета и других компонентов.

Исследователи полагают, что предложенные ими методики одинаково эффективны для поиска ботов любой категории, поскольку не выделяли различий среди бот-аккаунтов и не учитывали их в работе. Перспективы дальнейшего исследования проблемы мы видим в детальном изучении выделенных категорий социальных ботов.

6. Заключение

Результаты проведённого исследования корпуса текста постов аккаунтов, признанных социальными ботами, меняют сложившееся понимание социального бота. Разработанная модель социального бота на примере американского политического дискурса представляет основу для новой типологии ботов, при которой боты изучаются не только с точки зрения вида деятельности и назначения, но ещё и с точки зрения природы их существования. Выделение такого критерия оценки ботов позволит проводить направленные исследования для обнаружения конкретного типа бота.

Суть полученных результатов заключается в том, что обнаружение социального бота не должно сводиться к поиску результатов работы ПО, которое работает по некоторому алгоритму и воспроизводит заданные речевые паттерны с использованием некоторого ПО, в том числе с использованием технологии машинного обучения, к примеру, применяя для генерации текстов постов возможности ChatGPT. Необходимо комплексно подходить к задаче обнаружения социального бота, одним из этапов решения которой является определение природы социального бота и способа управления им.

Результативность известных методик обнаружения социальных ботов может быть переоценена при тестировании их работы на корпусах текстов социальных ботов, которые пред-

ставляют различные категории нарушителей. Так, например, методика, эффективная для поиска веб-ботов, может показать более низкие показатели результативности в случае, если бот-аккаунтом управляет человек.

Литература

1. Интернет в России: что говорит статистика? [Электронный ресурс]. URL: https://rskrf.ru/tips/eksperty-obyasnyayut/internet-stats/ (дата обращения: 20.11.2023).
2. *Ibtisam A. A.* Social bots' role in online political communication: Evidence from German Federal Election 2021 // MAGKS Joint Discussion Paper Series in Economics. 2023. № 09.
3. *Keller F. B., Schoch D., Stier S., Yang J. H.* Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign // Political Communication. 2020. № 37:2. P. 256–280.
4. *Ветров И.* Интернет победил телевизор. Как интернет и соцсети помогли Трампу победить Клинтон [Электронный ресурс]. URL: https://www.gazeta.ru/tech/2016/11/09/10318019/internetvstv.shtml (дата обращения: 26.03.2022).
5. *Почепцов Г.* Новая коммуникативная среда выборов и big data [Электронный ресурс]. URL: https://ms.detector.media/mediaanalitika/post/18419/2017-02-19-novaya-kommunykatyvnaya-sreda-vyborov-y-big-data/ (дата обращения: 26.03.2022).
6. Internet Archive [Электронный ресурс]. URL: https://web.archive.org/(дата обращения: 25.02.2022).
7. 3 times bots have impacted major world events [Электронный ресурс]. URL: https://www.netacea.com/blog/3-times-bots-have-impacted-major-world-events/?amp#2016_and_2020_The_US_presidential_elections_-_Did_bots_influence_the_result (дата обращения: 26.03.2022).
8. *Василькова В. В., Легостаева Н. И.* Социальные боты в политической коммуникации // Вестник Российского университета дружбы народов. Серия: Социология. 2019. Т. 19, № 1. С. 121–133.
9. *Василькова В. В., Легостаева Н. И., Радушевский В. Б.* Тематический ландшафт бот-пространства социальной сети «Вконтакте» // Журнал социологии и социальной антропологии. 2019. Т. 22, № 4. С. 202–245.
10. *Гудков А. С.* Анализ активности бот-аккаунтов в новостных сообществах социальной сети Вконтакте // Материалы 11-й Международной конференции «Управление развитием крупномасштабных систем» (MLSD), Москва, 1–3 октября 2018 года. Том II. С. 512–514.
11. *Ненашев С. М.* Информационно-технологическая и информационно-психологическая безопасность пользователей социальных сетей // Вопросы кибербезопасности. 2016. № 5 (18). С. 65–72.
12. *Логинова А. О.* Обнаружение интернет-бота по структурно-вероятностной модели электронного сообщения // Вестник Воронежского института МВД России. 2022. № 3. С. 105–114.
13. *Логинова А. О., Алейникова Д. В.* Выявление демаскирующих признаков социального бота на синтаксическом уровне генерируемого сообщения // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2023. № 1. С. 139–147.
14. Менщиков А. А. Методы обнаружения и противодействия автоматизированному сбору информации с веб-ресурсов: дис. … канд. тех. наук. СПб. 2019. 264 с.
15. Чесноков В. О. Алгоритмическое и программное обеспечение анализа графов ближайшего окружения для выявления ботов и определения неуказанных атрибутов пользователей в онлайновых социальных сетях: автореф. дис. … канд. М. 2018. 19 с.
16. *Василькова В. В. Легостаева Н. И.* Детектирование тематического разнообразия ботне-

тов: подходы и методика // Материалы XXV Международной конференции памяти профессора Л. Н. Когана «Культура, личность, общество в условиях пандемии и постпандемии: методология, опыт эмпирического исследования», УГПУ, 2022. С. 239–24.

Логинова Алина Олеговна
ведущий специалист по защите информации отдела информационной безопасности Департамента цифрового развития, Московский государственный лингвистический университет (ФГБОУ ВО МГЛУ, 119034, Москва, ул. Остоженка, д. 38, стр. 1), e-mail: a.loginova@linguanet.ru, ORCID ID: 0000-0002-7806-0586.

Автор прочитал и одобрил окончательный вариант рукописи.
Автор заявляет об отсутствии конфликта интересов.

The Descriptive Model of Social Bot on the Base of American Political Discourse

Alina O. Loginova

Moscow State Linguistic University (MSLU)

Abstract: The purpose of this study is to create a descriptive model of a social bot that takes into account its potential as a violator of information and psychological security. The study was made on the base of American political discourse. The model is based on the results of a comprehensive analysis of the unmarked text corpus of social bot posts on Twitter. The corpus of texts is a collection of texts of posts in English of social network accounts that were previously identified by researchers as social bot accounts involved in the president election campaign. The study is motivated by the lack of an exhaustive definition of the concept of "social bot". This bot model forms a new approach to the disclosure of the essence of this concept. In this article, a social bot is considered as a violator of the information and psychological security of the user of the Internet mass media. The article was funded by the Ministry of Digital Development of the Russian Federation according to the research project No. № 14/23-К.

Keywords: social bots, methods of detecting bots, information-psychological security, Internet mass media.

For citation: Loginova A. O. The descriptive model of Social Bot on the base of American political discourse (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 3, pp. 3-13. https://doi.org/10.55648/1998-6920-2024-18-3-3-13.

Content is available under the license Creative Commons Attribution 4.0 License

© Loginova A. O., 2024

The article was submitted: 22.01.2024;
revised version: 25.02.2022;
accepted for publication 27.02.2022.

References

1. *Internet v Rossii: chto govorit statistika?* [The Internet in Russia. What are the statistics about this?], available at: https://rskrf.ru/tips/eksperty-obyasnyayut/internet-stats/ (accessed 20.11.2023).
2. Ibtisam A. A. Social bots' role in online political communication: Evidence from German Federal Election 2021. *MAGKS Joint Discussion Paper Series in Economics*, 2023, no. 09.
3. Keller F. B., Schoch D., Stier S., Yang J. H. Political Astroturfing on Twitter: How to Coordinate a Dis-

information Campaign. *Political Communication*, 2020, no. 37:2. pp. 256-280.
4. Vetrov I. *Internet pobedil televizor. Kak internet i socseti pomogli Trampu pobedit' Klinton* [The Internet has won over the TV. How the Internet and social media helped Trump defeat Clinton], available at: https://www.gazeta.ru/tech/2016/11/09/10318019/internetvstv.shtml (accessed: 26.03.2022).
5. Pochepcov G. *Novaya kommunikativnaya sreda vyborov i big data* [The new communicative environment of elections and big data], available at: https://ms.detector.media/mediaanalitika/post/18419/2017-02-19-novaya-kommunykatyvnaya-sreda-vyborov-y-big-data/ (accessed: 26.03.2022).
6. *Internet Archive*, available at: https://web.archive.org/ (accessed: 25.02.2022).
7. *3 times bots have impacted major world event*s, available at: https://www.netacea.com/blog/3-times-bots-have-impacted-major-world-events/?amp#2016_and_2020_The_US_presidential_elections_-_Did_bots_influence_the_result (accessed: 26.03.2022).
8. Vasil'kova V. V., Legostaeva N. I. Social'nye boty v politicheskoj kommunikacii [Social bots in political communication]. *Vestnik Rossijskogo universiteta druzhby narodov. Seriya: Sociologiya,* 2019, vol. 19, no. 1. pp. 121-133.
9. Vasil'kova V. V., Legostaeva N. I., Radushevskij V. B Tematicheskij landshaft bot-prostranstva social'noj seti «Vkontakte» [Thematic landscape of the bot-space of the social network “Vkontakte”]. *ZHurnal sociologii i social'noj antropologii,* 2019, vol 22, no. 4. pp. 202-245.
10. Gudkov A. S. Analiz aktivnosti bot-akkauntov v novostnyh soobshchestvah social'noj seti Vkontakte [The analysis of bot-accounts activity in news communities of the Vkontakte social network]. *Materialy odinnadcatoj mezhdunarodnoj konferencii,* vol. II, 2018, pp.512-514.
11. Nenashev S. M. Informacionno-tekhnologicheskaya i informacionno-psihologicheskaya bezopasnost' pol'zovatelej social'nyh setej [Information-technical and information-psychological security of social-network users]. *Voprosy kiberbezopasnosti,* 2016, no. 5 (18), pp. 65-72.
12. Loginova A. O. Obnaruzhenie internet-bota po strukturno-veroyatnostnoj modeli elektronnogo soobshcheniya [Detecting internet bot by a structural probabilistic model of its electronic message]. *Vestnik Voronezhskogo instituta MVD Rossii*, 2022, no. 3, pp. 105-114.
13. Loginova A. O., Alejnikova D. V. Vyyavlenie demaskiruyushchih priznakov social'nogo bota na sintaksicheskom urovne generiruemogo soobshcheniya [Detecting the unmasking features of a social bot at the syntax level of a generated message]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyj analiz i informacionnye tekhnologii*, 2023, no 1, pp. 139-147.
14. Menshchikov A. A. *Metody obnaruzheniya i protivodejstviya avtomatizirovannomu sboru informacii s veb-resursov* [Methods of detecting and withstanding automated collection of information from web resources]. Ph. D. thesis. Saint Petersburg, 2019. 264 p.
15. CHesnokov V. O. *Algoritmicheskoe i programmnoe obespechenie analiza grafov blizhajshego okruzheniya dlya vyyavleniya botov i opredeleniya neukazannyh atributov pol'zovatelej v onlajnovyh social'nyh setyah* [Detecting the thematic diversity of botnets: approaches and methods]. Ph. D. thesis. Moscow, 2018. 19 p.
16. Vasil'kova V. V. Legostaeva N. I. Detektirovanie tematicheskogo raznoobraziya botnetov: podhody i metodika [Detecting the thematic diversity of botnets: approaches and methods]. *Kul'tura, lichnost', obshchestvo v usloviyah pandemii i post-pandemii: metodologiya, opyt empiricheskogo issledovaniya. Materialy XHV Mezhdunarodnoj konferencii pamyati professora L. N. Kogana*, Ekaterinburg, Ural State Pedagogical Univercity, 2022, pp. 239-24.

Alina O. Loginova

Leading specialist of the Information Security Sector of the Department of Digital Development, Moscow State Linguistic University (MSLU, 119034, Moscow, Ostozhenka street, 38, build. 1), e-mail: a.loginova@linguanet.ru, ORCID ID: 0000-0002-7806-0586.

DOI: 10.55648/1998-6920-2024-18-3-14-27 УДК 004.9

Методика формирования проектных команд на основе ролевой модели Белбина

Е. В. Вершинина, О. М. Зверева, С. Н. Шабунин

ФГАОУ ВО «УрФУ имени первого Президента России Б. Н. Ельцина»

Аннотация: В данной статье представлена методика формирования команд, способных успешно осуществлять проектную деятельность. Обычно при формировании проектных команд учитывается лишь функциональная составляющая, т.е. наличие у членов команды профессиональных знаний и навыков, необходимых для создания определенного продукта. Опыт показывает, что это еще не гарантирует успех. С другой стороны, существуют методики определения индивидуальных особенностей людей, но они, как правило, замкнуты в рамках теории и практики менеджмента и направлены на формирование команд специалистов какого-то определенного профиля, не учитывая различия в профессиональных компетенциях. Отличительной особенностью предложенной методики является сочетание учета профессиональных компетенций и навыков, определяемых как специализации, и индивидуальных психологических особенностей потенциальных участников команды. При этом индивидуальные особенности предлагается определять на основе ролевой модели Белбина – модели, ориентированной на успешное выполнение проектов. Созданное приложение, используя хранящиеся данные о кандидатах, формирует все возможные по составу команды. Далее для каждой из команд рассчитывается показатель разнообразия ролей в соответствии с моделью Белбина. Именно этот показатель может служить оценкой потенциальной эффективности команды. В результате рекомендуется для выполнения проекта та команда, у которой показатель разнообразия ролей имеет максимальное значение. Наличие участника, обладающего качествами лидера, также является важным фактором успеха, поэтому приложение проверяет и выдает сообщение о том, есть ли кандидат с данной ролью в сформированной команде. Приложение, автоматизирующее данную методику, позволяет просматривать данные о членах команды и предоставляет возможность редактировать ее состав. Предложенная методика может быть применена как в компаниях, где существует возможность выбора из нескольких работников с набором схожих профессиональных компетенций, так и в учебном процессе, где командная работа над проектом становится одной из основных форм самостоятельной работы студентов и где также доступны разные варианты построения команд. Она прошла успешное тестирование в одной из ИТ-компаний уральского региона, а также доказала свою перспективность при формировании учебных команд в Департаменте информационных технологий и автоматики и на кафедре радиоэлектроники и телекоммуникаций Института радиоэлектроники и информационных технологий – РТФ Уральского федерального университета имени первого Президента России Б. Н. Ельцина.

Ключевые слова: команда, веб-приложение, проект, жизненный цикл, модель Белбина, учебный процесс, программный инструмент.

Для цитирования: Вершинина Е. В., Зверева О. М., Шабунин С. Н. Методика формирования проектных команд на основе ролевой модели Белбина // Вестник СибГУТИ. 2024. Т. 18, № 3. С. 14–27. https://doi.org/10.55648/1998-6920-2024-18-3-14-27.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Вершинина Е. В., Зверева О. М., Шабунин С. Н., 2024

Статья поступила в редакцию 17.08.2023;
переработанный вариант – 06.12.2023;
принята к публикации 22.02.2024.

1. Введение

В реалиях сегодняшнего дня множество работ выполняется в форме проекта. Сфера информационных технологий, с которой начиналось данное исследование, – не исключение: можно считать, что она – одна из сфер, где форма проекта – основная. Любой информационный проект – это комплексная задача, которую способна выполнить только команда профессионалов. Создать эффективно работающий коллектив, способный успешно выполнить проект – задача нетривиальная, т.к. необходимо учитывать целый ряд показателей, которые часто невозможно оценить численно, – речь идет о ментальных, психологических и коммуникационных характеристиках участников проекта.

Один из базовых принципов рассматриваемой методики основан на понятии жизненного цикла (ЖЦ) продукта, который должен стать результатом выполнения проекта. Если рассматривать информационную отрасль как сферу применения, то таким продуктом является программное обеспечение (ПО). На практике в большинстве фирм-производителей ПО именно функциональный принцип, отражающий этапы разработки как части ЖЦ ПО, заложен в основу разделения труда среди сотрудников. В соответствии с этим принципом определяются специализации сотрудников и именно на нем основано формирование команд.

Если фирма небольшая, то набор сотрудников в проектную команду очевиден – в нее включаются свободные от выполнения других проектов сотрудники, обладающие в сумме специализациями, достаточными для выполнения проекта. В больших и средних компаниях в условиях возможности выбора встает вопрос: по какому принципу выбрать сотрудника из множества имеющих одинаковую специализацию?

Подобная проблема возникает и в образовательном процессе, когда академическую группу или целый поток студентов нужно разбить на команды для выполнения учебных проектов. Часто такие команды создаются на основе личностных симпатий или дружественных связей. Вопрос в том, всегда ли команды, сформированные по такому принципу, будут успешными?

Рассматриваемая методика, будучи основанной на функциональном принципе формирования команд, предлагает учитывать индивидуальные особенности кандидатов, которые определяются в соответствии с ролевой моделью Белбина (R. M. Belbin) [1]. В качестве оценки потенциальной эффективности команды предлагается использовать показатель разнообразия ролей, который считается как количество различных доминантных ролей по модели Белбина, диагностированных у участников команды. В качестве оптимального состава команды предлагается выбрать тот состав, который имеет максимальное значение данного показателя.

2. Проект и команда: основные определения

Введем основные понятия и концепции, которые будут использованы далее.

Согласно Руководству PMBOK [2]: «проект – это временное предприятие, направленное на создание уникального продукта, услуги или результата». К основным характеристикам проекта можно отнести наличие временных рамок, уникальность результата и ограниченность ресурсов. Главным отличием проекта от процесса является то, что проект выполняется однократно и в результате работы над проектом создается уникальный продукт, а процесс – это набор процедур, которые могут быть повторены через некоторое время. Для выполнения проекта создается проектная команда. В рассматриваемой постановке задачи команда создается для выполнения определенного проекта.

Введем понятие «команда». В 1993 г. Jon R. Katzenbach и Douglas K. Smith изложили своё понимание команды в книге, название которой на русский язык было переведено как «Командный подход: Создание высокоэффективной организации» (хотя при точном переводе оно выглядит как «Мудрость команд: создание высокоэффективной организации») [3]. Согласно предложенной концепции, которая считается классической, «команда проекта – это

малочисленная группа людей, которые владеют необходимыми для достижения единой цели знаниями и умениями и совместно отвечают за достижение результата» [3].

Приведем еще один пример определения команды, процитировав российского ученого – профессора Р. Л. Кричевского, который определяет команду следующим образом: «Команда – особая разновидность малой группы …. Это небольшая группа людей (возможный объём от 2 до 25 человек, но обычно не более 10–12 человек), которые:

– взаимодействуют и взаимодополняют друг друга;
– преследуют значимые общие цели, разделяя ответственность за их достижение;
– связаны единым замыслом;
– продуманно позиционируют себя для выполнения специфических ролей или функций, демонстрируя способность к быстрому и гибкому реагированию на изменение среды;
– отличаются зрелым профессионализмом и стремятся к высокому стандарту индивидуального и коллективного выполнения» [4].

Из этих определений следует, что для успешной реализации проекта команда должна состоять из профессионалов, суммарно обладающих специальными навыками и знаниями, обеспечивающими поддержание всех этапов жизненного цикла создаваемого продукта. Можно назвать это функциональной составляющей команды, т.к. речь идет о выполнении определенных функций.

Все приведенные определения команд показывают, что наличия полного набора профессиональных компетенций недостаточно для успешной работы команды. Требуются дополнительные факторы, такие как: единые цели, единый замысел и т.д. Эти характеристики сложно оценить, т.к. они зависят от многих причин, в том числе внешних по отношению к команде. В определении, введенном в [4], утверждается необходимость взаимодополнения членами команды друг друга и выполнения ими специфических ролей. Именно эти дополнительные характеристики призвана учесть описываемая методика. Предлагается для реализации этой задачи использовать ролевую модель Белбина.

3. Особенности определения функциональной составляющей для учебных проектов

Набор специализаций, необходимых для выполнения проекта, чаще всего известен заранее. В компаниях, которые существуют на рынке, накоплен опыт выполнения проектов в определенной сфере, и они могут задать такой набор до начала выполнения нового проекта.

Определенное своеобразие имеют учебные проекты. Одним из приложений методики стало ее использование в учебном процессе. Первый опыт такого применения был в департаменте информационных технологий и автоматики, где студентами выполняются информационные проекты. Здесь проявилась определенная специфика – необходимо было выделить некоторые «универсальные» специализации, с помощью которых можно было бы выполнить большинство таких проектов. Обращение к опыту ИТ-компаний не решило проблемы из-за особенностей учебного процесса, т.к. нужен был некий стандарт на все проекты, а компании имели возможность более тонкой настройки, поэтому специализации от проекта к проекту могли отличаться, как отличались и названия специализаций внутри разных компаний.

Было решено набор специализаций определить через понятие ЖЦ программного продукта. Одним из первых этот цикл описал Winston W. Royce в своей статье «Managing the Development of Large Software Systems» («Управление разработкой больших программных систем») [5], он показан на рис. 1.

Рис. 1. Жизненный цикл ПО по У. Ройсу

Описанная У. Ройсом модель ЖЦ получила название модели «водопада», она состоит из 7 этапов: определение системных требований, определение требований к ПО, анализ, проектирование, кодирование, тестирование, эксплуатация. Существуют и другие модели ЖЦ ПО [6, 7], они различаются структурой, повторяемостью этапов, но сами этапы имеют сходное содержание.

Теме ЖЦ ПО посвящены и отраслевые стандарты, такие как ГОСТ 34.601-90 [8], ГОСТ Р ИСО/МЭК 12207-2010 [9]. К их особенностям можно отнести то, что первый – связывает стадии ЖЦ с созданием документации на программный продукт, второй – ориентирован на процессный подход.

При формировании команды для учебного проекта необходимо было также учитывать тот факт, что в команде не должно быть больше 5–6 человек, иначе сложно оценивать вклад каждого студента.

Были выделены следующие специализации:

– Аналитик: основное участие принимает на этапе формирования системных требований, требований к ПО и анализа. Основной функционал – сбор и анализ информации, связанной с проектом, формирование планов для работы над проектом. В реальных производственных условиях перечень функций аналитика определяется способом организации производства, квалификацией специалиста, спецификой моделируемой предметной области.

– Разработчик: основное участие принимает на этапе проектирования и кодирования. Основной функционал – проектирование системы, написание кода.

– Дизайнер: основное участие принимает на этапе проектирования и кодирования. Основной функционал – проектирование и создание визуальной части программного продукта. Отвечает за внешний вид и восприятие информации, разрабатывает удобный интерфейс, представление информации на экране.

– Тестировщик: основное участие принимает на этапе тестирования. Основной функционал – проведение всех видов тестирования, в том числе автоматизированного тестирования (создание своих тестов), написание тест-планов, выдача рекомендаций по исправлению ошибок или их самостоятельное исправление.

Кроме этих четырех специализаций, в связи с необходимостью централизации управления и взаимодействия с куратором проекта была введена пятая специализация – лидер команды, чаще называемый сокращенно «тимлид». Тимлид участвует во всем цикле проекта, на всех его стадиях. Его/её основной функционал состоит в том, что он/она отвечает за выполнение планов на уровне команды, организацию её работы, развитие технической части всего проекта в соответствии с принятыми правилами.

Для учебных проектов в сфере радиоэлектроники был предложен следующий набор специализаций: схемотехник, трассировщик печатных плат, программист микроконтроллеров, конструктор, специалист в области электромагнитной совместимости, антенн и высокочастотных устройств.

В общем случае на первом этапе нужно выявить (или заявить) список специализаций, которые должны присутствовать в команде. Понятие роли используется на следующих этапах и вводится в следующем подразделе.

4. Ролевые модели

4.1. Опыт использования ролевых моделей

Практика показывает, что наличие в команде специалистов с нужными специализациями, обеспечивающими набор необходимых знаний и практических навыков, не всегда обеспечивает эффективную работу, а значит, и успех команды в целом. Команда состоит из людей, каждый из которых обладает своими психологическими и ментальными особенностями, «мягкими» (soft) навыками, которые не могут не повлиять на ход проекта и на его результат.

Не существует единого мнения по вопросу о том, какая команда наиболее успешна: состоящая из похожих по психологическим и ментальным особенностям людей (гомогенная) или из отличающихся друг от друга (гетерогенная). Однако в последнее время все больше исследователей склоняется в сторону гипотезы о гетерогенности. Об этом свидетельствуют такие работы, как: L. Dyaram и T. J. Kamalanabhan [10], название статьи переводится на русский язык как «Разнообразие в командах разработчиков программного обеспечения: управляемые эффекты взаимонезависимости, срока нахождения в команде и типа проекта» или статья авторов: B. F. Hsu, W. L. Wu, R. S. Ye [11], название которой переводится как «Личностный состав команды, аффективные связи и обмен знаниями: анализ на уровне команды». В других же исследованиях [12, 13] авторы предупреждают о возникающих в таких случаях конфликтных ситуациях из-за разницы культур и недопонимании при общении.

С концепцией гетерогенности связана идея командных ролевых моделей. Она заключается в том, что если в команде ее члены выполняют определенный набор ролей, то такая команда способна эффективно работать.

Под ролью в данном случае понимается паттерн поведения, характеризующий способ взаимодействия одного члена команды с другим, где его действия способствуют прогрессу команды в целом. Количество ролей в моделях варьируется от четырех – в модели, описанной в книге G. M. Parker [14], автора многих книг по командной работе, до пятнадцати, такая модель описана в книге J. Davis, P. Millburn, T. Murphy, M. Woodhouse, названной «Успешный тимбилдинг: как создать действительно работающую команду» [15].

Одной из наиболее известных ролевых моделей команд является модель Белбина [1]. Успешное её применение описано в статьях: J. S. Prichard, N. A. Stanton «Тестирование командно-ролевой модели эффективных групп Белбина» [16], Park W.-W., Bang H. «Баланс ролей в команде и производительность команды» [17] и C. Zarzu, C. Scarlat, P. Falcıoğlu «Состав команды и ее эффективность: достижение более качественных результатов в международной среде высшего образования» [18]. Один из авторов данной статьи имеет положительный опыт применения этой модели в рамках учебного процесса, который более детально изложен в [19]. В этой работе подробно описан случай, когда сформированные на основе дру-

жеских связей команды выполняли учебные проекты, затем эти команды были протестированы с использованием опросника модели Белбина, и полученный результат на практике подтвердил тот факт, что самые успешные команды содержали наибольшее число различных ролей по Белбину.

Для объективности следует привести и те работы, в которых ценность данной модели подвергается сомнению. Это, например, статья D. Partington, H. Harris «Баланс ролей в команде и эффективность команды: экспериментальное исследование» [20], где данная модель критикуется за отсутствие ясного теоретического обоснования.

К этому же множеству можно отнести статью с названием, в котором уже звучит это сомнение: «Роли в команде и производительность: есть ли «действительно» связь?» [21]. В этой статье описан эксперимент с применением рассматриваемой модели при формировании одиннадцати команд, при этом только в четырех из них данная зависимость («связь» в терминах статьи) была обнаружена.

4.2. Модель Белбина

Приведем краткое описание модели Белбина. Все роли этой модели можно условно разделить на 3 группы: ориентированные на решение проблем (интеллектуальные), ориентированные на взаимодействие с людьми (социальные) и ориентированные на действие.

Интеллектуальные роли:

– Генератор идей (Plant): изобретатель и новатор, креативен, способен находить нетрадиционные пути для решения проблем;

– Аналитик-стратег/оценщик (Monitor Evaluator): объективный и проницательный; анализирует возможности выполнения проекта, редко ошибается в оценке ситуации; перед принятием решений тщательно анализирует ситуацию, рассматривая ее с разных сторон.

Социальные роли:

– Координатор (Coordinator): обычно это лидер команды; распределяет задачи, устанавливает сроки выполнения, контролирует результат; хорошо знает свою команду, использует знания и навыки каждого сотрудника;

– Разведчик / Исследователь ресурсов (Resource Investigator): основная задача в команде – получение новых идей, общителен, генерирует новые идеи, собирая и преобразовывая идеи других членов команды; используя навыки коммуникации, создает сеть знакомств, а затем получает через нее данные и ресурсы;

– Душа команды/ Коллективист (Team Worker): мягкий и дипломатичный, в команде отвечает за урегулирование внутренних конфликтов.

Роли действия:

– Мотиватор (Shaper): энергичный и ориентированный на успех; такой сотрудник сам обладает высоким уровнем мотивации и мотивирует других.

– Реализатор (Implementer): дисциплинированный и трудолюбивый исполнитель; при выполнении задач всегда руководствуется здравым смыслом;

– Контролер (Completer/Finisher): наиболее эффективен в конце проекта, чтобы завершить проект и тщательно проверить работу на наличие ошибок; умеет составлять документацию.

Более подробно о каждой роли с обсуждением ее психологических особенностей, сильных сторон и недостатков можно узнать из материалов, представленных на сайте HR-академии [22]. Полезные материалы с описанием реальных ситуаций, которые возникают, когда часть ролей не представлена, приведены в статье М. Цепкова «Типология ролей в команде» [23]. Известный сайт «Хабр» [24] содержит материалы с описанием методов «лечения», когда ролей не хватает или они исполняются неправильно.

Каждый человек в силу своих психологических особенностей и воспитания склонен к выполнению какой-то определенной роли из тех, что представлены в модели Белбина. Для определения этой роли предлагается заполнить специальный опросник (Self-Perception Inven-

tory), который является центральным звеном данной модели. Опросник состоит из 8 групп высказываний по 8 в каждой группе. В каждой группе респонденту необходимо распределить 10 баллов, оценив важность этих высказываний для себя. По заполненному опроснику подсчитывается (в баллах) склонность респондента к выполнению той или иной роли. Чем выше набранный по некоторой роли балл, тем более вероятно, что в команде данный сотрудник будет выполнять именно эту роль, т.к. она для него более комфортна. Эту роль обычно называют доминантной.

Ролевая модель Белбина в своем «классическом» варианте разработана для членов команд, выполняющих сходные функции, т.е. в ней не учитываются профессиональные компетенции участников. В большинстве современных проектов нельзя не учитывать эти специализации, поэтому при формировании команды в данной методике предлагается основываться именно на профессиональных навыках. Кроме того, эта модель ориентирована на формирование команд, в которых не менее 8 человек (по количеству ролей), и в ней оценивается полнота реализации ролей: т.е. сколько из 8 ролей могут комфортно выполняться членами сформированной команды. Предлагаемая методика допускает формирование и меньших по составу команд, что обычно требуется в учебном процессе, поэтому был введен термин «разнообразие» ролей. Разнообразие ролей подразумевает количество разных ролей, которые являются доминантными для членов команды. В соответствии с описываемой методикой из всех возможных команд предлагается к выбору та, у которой этот показатель разнообразия ролей максимальный. В связи с тем, что методика автоматизирована, в приложении реализован полный перебор всех возможных по составу команд (состав определяется набором заданных специализаций), подсчет показателя разнообразия ролей для каждой из команд и выявление команд с максимальным значением показателя разнообразия ролей.

Как можно понять из вышеприведенного описания, опросник имеет нестандартную форму: нет привычных тестовых вопросов закрытого типа, предлагающих выбор одного или нескольких ответов из предложенных вариантов, или тестовых вопросов открытого типа, предлагающих ввод ответа в некоторое поле. В связи с этим готовую тестовую систему использовать не представлялось возможным, и при создании приложения был разработан отдельный тестовый модуль. Окно экрана с работающим тестовым модулем показано на рис. 2.

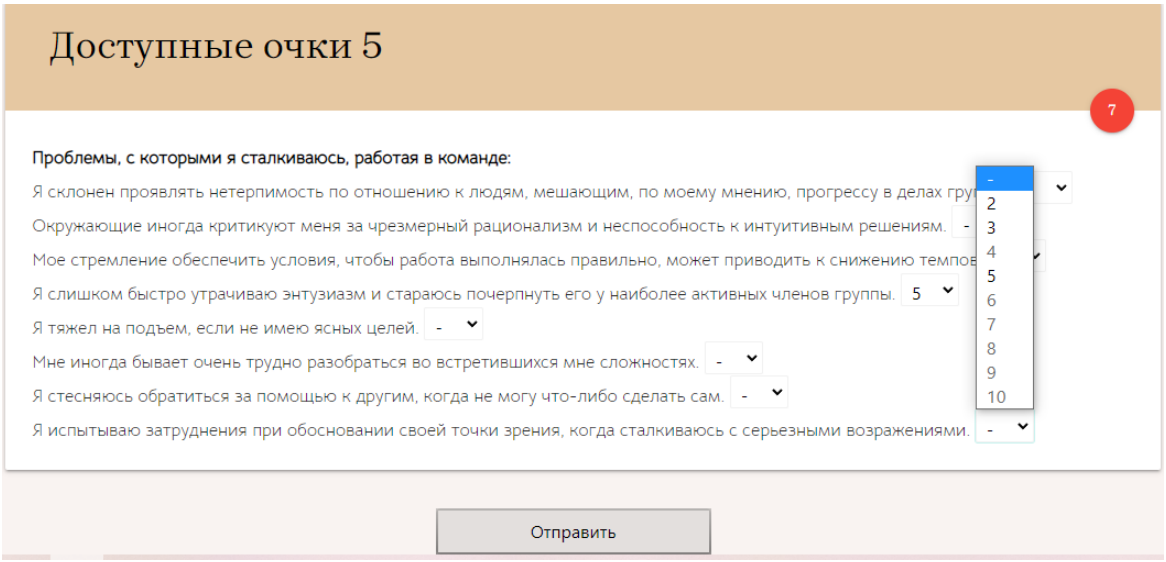


Рис. 2. Вид экрана с опросом

Из рис. 2 видно, что в текущий момент респондент уже выделил 5 баллов из 10, и ему/ей осталось распределить еще 5 баллов (очков). Это является подсказкой для проходящего опрос.

Одним из этапов методики является тестирование кандидатов с целью определения доминантных ролей для всех претендентов на участие в проекте. Это тестирование необходимо провести до начала формирования команды.

5. Реализация методики

Для автоматизации процесса формирования команды на основе предлагаемой методики было создано веб-приложение.

В приложении реализовано две роли: можно авторизоваться и работать как Сотрудник или как Руководитель. Эти пользователи имеют разные права в системе, т.е. набор возможных действий в системе. Формированием (созданием) команды занимается Руководитель. Стартовый экран Руководителя показан на рис. 3.

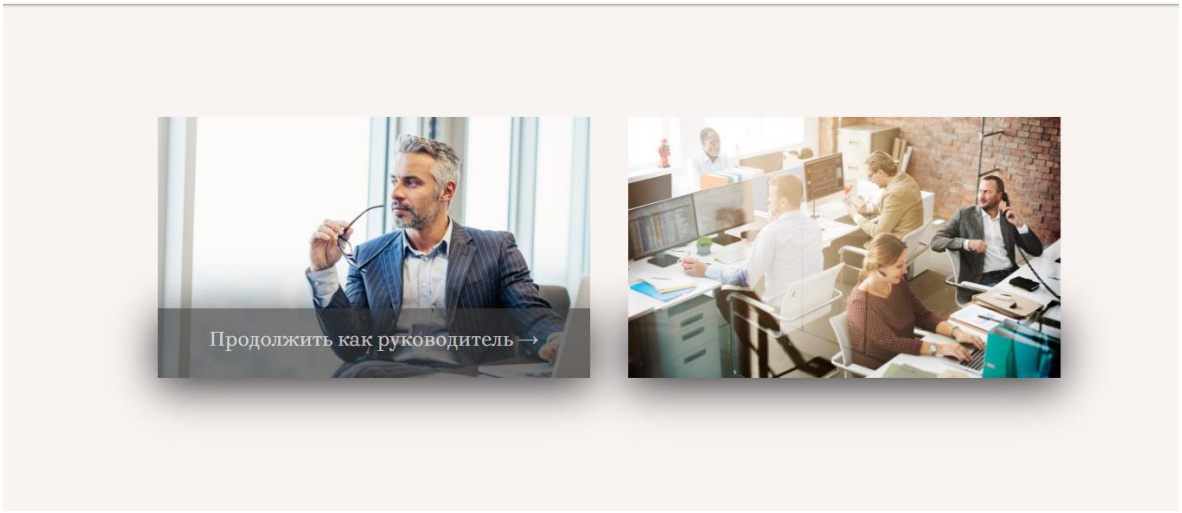


Рис. 3. Стартовый экран Руководителя

Руководитель с целью формирования команды для определенного проекта запускает модуль «Создание проектной команды». В ходе работы:

1) формируется карточка проекта, в которой обязательно заполняется набор необходимых специализаций;

2) из незанятых на других проектах сотрудников (занятость определяется через обращение к встроенной БД) формируются все возможные варианты команд, каждая команда содержит набор заданных специализаций;

3) для каждой команды вычисляется показатель «Разнообразие ролей», который записывается в формате:

Число_различных_ролей_в_команде/Общее_число_разных_ролей;

4) на экран выводится команда с максимальным значением оценки «Разнообразие_ролей». Если таких команд несколько, можно в дальнейшем просмотреть все эти команды;

5) для каждой из выведенных команд определяется, есть ли лидер (сотрудник, имеющий доминантную роль «Координатор»), результат выводится на экранную форму (показана на рис. 4);

6) предоставляется возможность редактирования команды: можно произвести замену любого из членов команды, показатели будут пересчитаны;

7) результат в виде сформированной команды сохраняется под определенным именем; сотрудники, включенные в команду, в БД будут помечены как «занятые в проекте».

На рис. 4 показан экран работающего приложения. Была заявлена команда из 4 специализаций: разработчик, дизайнер, аналитик, тестировщик. В данном примере не был заявлен тимлид (такая специализация типична для учебных команд), однако в компании, для которой разработана первая версия приложения, понимая значение такого члена в команде, заявили требование о необходимости вывода сведений о наличии в сформированной

команде сотрудника с лидерскими качествами. Лидером считается член команды, доминантная роль которого по модели Белбина «Координатор».

Рис. 4. Вид экрана со сформированной командой

Приложение выдает на экран (рис. 4) сформированный состав команды с максимальным показателем разнообразия ролей. При этом указано, что лидер в команде есть. Если лидера в команде нет, Руководитель может принять решение о вводе или замене одного из участников, введя сотрудника с доминантной ролью Координатора (т.е. произвести редактирование состава команды). После редактирования команды эти сотрудники в БД приложения помечаются как занятые в определенном проекте.

Формирование учебных команд производится циклически с участием пользователя, выполняющего роль Руководителя. Цикл выполняется до тех пор, пока количество не занятых в проекте студентов не станет меньше числа специализаций, заявленных для команды.

Оставшиеся студенты могут быть добавлены в сформированные команды или по их согласию можно сформировать отдельную команду, или выбрать какой-то другой устраивающий всех вариант распределения.

В процесс формирования команд в данный момент вовлечен пользователь, он проходит в автоматизированном режиме. Предполагается, что вторая версия приложения будет выполнять этот цикл автоматически.

Дать численную оценку эффективности предложенной методики формирования проектных команд достаточно сложно. Это объясняется тем, что ряд проектов выполняется в течение одного семестра, другие для своего успешного завершения требуют несколько семестров. Длительность выполнения зависит от заказчиков проекта и его сложности. Вычисление оценки эффективности на основе баллов, полученных за проект, тоже вызывает определенные сомнения, т.к. они выставляются комиссиями с разным составом, что не может не внести ошибку в результирующую оценку эффективности.

Для компаний с развитой проектной деятельностью предложенная методика и разработанное на её основе приложение должны быть привлекательными с точки зрения автоматизации процеса формирования проектных команд и контроля текущей занятости сотрудников. При этом команда, сформированная на основе предложенного подхода, будет иметь достаточно высокий профессиональный потенциал, также будут учтены психологические и ментальные особенности её членов.

О положительном эффекте использования методики и приложения в учебном процессе можно говорить в том числе в свете облегчения работы организаторов проектных

практикумов на этапе формирования команд. Практика использования методики показывает увеличение доли положительных оценок, полученных командами от внешних заказчиков проектов.

6. Заключение

Предложена методика формирования команд для выполнения проектов. Наряду с набором специализаций, которые необходимы для успешной реализации проекта в профессиональном плане, учитываются индивидуальные особенности её будущих членов, что закладывает потенциал для эффективной работы.

Методика состоит из следующих этапов:

– ввод информации о компании и сотрудниках;

– тестирование сотрудников для определения доминантных ролей по модели Белбина;

– ввод информации о проекте, в том числе о количестве и наборе специализаций сотрудников, необходимых для выполнения проекта;

– формирование команд с выполнением условий функциональной составляющей и учетом ролей по модели Белбина;

– редактирование команды в соответствии с политикой компании или по другим причинам.

Методика и первая версия приложения прошли тестирование в одной из информационных компаний Уральского региона, получив положительную оценку.

С введением в образовательные программы УрФУ модуля «Проектная деятельность», в рамках которого создаются студенческие команды для выполнения проектов по заявкам внешних организаций, предложенная методика также была успешно опробована и в учебном процессе. Основной сферой приложения методики стала область ИТ-проектов и формирования команд для их реализации. Также был получен первый положительный опыт применения методики и приложения при реализации магистерской программы «Инженерия радиоэлектронных средств и систем».

Приложение имеет модульную структуру, что обеспечивает его легкую расширяемость. Возможно расширение приложения в части тестирования (т.к. существуют и другие ролевые модели, признанные специалистами) за счет подключения других тестовых модулей. Планируется выпуск второй версии программного продукта, ориентированного на использование в учебном процессе.

Следует отметить, что описанная методика не гарантирует успех командной работы, на который, безусловно, влияют и другие факторы, в том числе: технологии, на основе которых будет выполняться проект, опыт и квалификация членов команды, лидерские качества руководителя команды, нельзя исключать и влияние различных внешних факторов. Можно с уверенностью утверждать только то, что, при прочих равных условиях, команды, сформированные на основе предложенной методики, имеют более высокие шансы на успех.

Литература

1. *Belbin R. M.* Team Roles at Work. Routledge, 2010. 162 p.
2. Руководство к своду знаний по управлению проектом (Руководство PMBOK): Институт управления проектами, Project Management Institute, Inc, 2017. 762 с.
3. *Katzenbach J. R., Smith D. K.* The Wisdom of Teams: Creating the High-Performance Organization. Harvard Business Review Press; Reprint edition (October 13, 2015). 304 p.
4. *Кричевский Р. Л.* Психология лидерства. М.: Стаут, 2007. 541 с.
5. *Winston W. Royce.* Managing the Development of Large Software Systems // Technical Papers of Western Electronic Show and Convention (WesCon), 1970. [Электронный ресурс]. URL:

http://leadinganswers.typepad.com/leading_answers/files/original_waterfall_paper_winston_royce.pdf (дата обращения: 22.07.2023).
6. *Сенник Ю. С., Гребенников Р. И.* Жизненный цикл информационных систем // Системный анализ и прикладная информатика. 2015. № 2. С. 4–9.
7. *Вольфсон Б.* Гибкие методологии разработки [Электронный ресурс]. URL: https://strategium.space/wp-content/uploads/2018/07/Gibkie-metodologii.pdf (дата обращения: 11.07.2023).
8. ГОСТ Р 59793-2021. Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Стадии создания. М.: Российский институт стандартизации, 2021. 6 с.
9. ГОСТ Р ИСО/МЭК 12207-2010. Информационная технология. Системная и программная инженерия. Процессы жизненного цикла программных средств. М.: Стандартинформ, 2011. 100 с.
10. *Dyaram L. and Kamalanabhan T. J.* Diversity in software development teams: the moderating effects of interdependence, team tenure and project type // International Journal of Business Innovation and Research. 2011. V. 5, № 2. P. 159–178.
11. *Hsu B. F., Wu W. L., Ye R. S.* Team personality composition, affective ties and knowledge sharing: a team-level analysis // International Journal of Technology Management. 2011. V. 53, № 2–4. P. 331–351.
12. *He Z. and Thatchenkery T.* Understanding culture-specific leadership relationship in a multicultural virtual project team (MVPT): a case study // International Journal of Business and Globalisation. 2011. V. 7, № 4. P. 446–458.
13. *Lech M. B., Georgiade T. J., Sukup J. E., Tomovic C. L.* The impact of Product Lifecycle Management on virtual team creation: a concept change model for heightened productivity // Int. Journal of Manufacturing Technology and Management. 2020. V. 19, № 3–4. P. 279–293.
14. *Parker G. M.* Team Players and Teamwork: The New Competitive Business Strategy. Wiley, 1990. 178 p.
15. *Davis J., Millburn P., Murphy T., Woodhouse M.* Successful Team Building: How to Create Teams that Really Work. Kogan Page, 1992. 184 p.
16. *Prichard J. S. and Stanton N. A.* Testing Belbin's team role theory of effective groups // Journal of Management Development. 1999. V. 18, № 8. P. 652–665.
17. *Park W.-W. and Bang H.* Team role balance and team performance // Proc. Belbin's Biennial Conf. "Changing Role of Management in the 21st Century", Cambridge, Clare College, 2002.
18. *Zarzu C., Scarlat C., Falcıoğlu P.* Team Composition and Team Performance: Achieving Higher Quality Results in an International Higher Education Environment // Proc. Int. Conf. "Active Citizenship by Knowledge Management & Innovation", June 19–21, 2013, Zadar, Croatia. P. 1321–1328.
19. *Zvereva O. M., Shams R.* Software Support for Team Engineering: Educational Case for IT Students // Proc. IV International Conference on Information Technologies in Engineering Education (Inforino), October 23–26, 2018, Moscow, Russia. P. 113–118. [Электронный ресурс]. URL: https://nrl.northumbria.ac.uk/id/eprint/40337/1/AAM_Zvereva%26Shams_2018%20Inforino%20Conference_Software%20Support%20for%20Team%20Engineering.pdf (дата обращения: 17.02.2024).
20. *Partington D. and Harris H.* Team role balance and team performance: an empirical study // Journal of Management Development. 1999. V. 18, № 8. P. 694–705.
21. *Senior B.* Team roles and team performance: Is there 'really' a link? // Journal of Occupational and Organizational Psychology. 1997. V. 70. P. 241–258.
22. В поисках команды: Белбин в кратком изложении с дополнением. HR-академия [Электронный ресурс]. URL: https://hr-academy.ru/hrarticle/v-poiskah-komandyi--belbin-v-kratkom-izlojenii-s-dopolneniem.html (дата обращения: 17.07.2023).

23. *Цепков М.* Типология ролей в команде. Mytype. Выпуск 2. Типологии. 1-2.2016 [Электронный ресурс]. URL: https://mtsepkov.org/Типология_ролей_в_команде_(статья_в_ _MyType_1-2.2016) (дата обращения: 18.07.2023).
24. Модель Белбина для IT: сила и слабость разных команд: Хабр, 2020. [Электронный ресурс]. URL: https://habr.com/ru/company/oleg-bunin/blog/522586/ (дата обращения: 22.07.2023).

Вершинина Екатерина Викторовна
студентка, ФГАОУ ВО Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, e-mail: Ek_Vershinina@mail.ru, ORCID ID: 0009-0008-1690-9143.

Зверева Ольга Михайловна
к.т.н., доцент департамента информационных технологий и автоматики, ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б. Н. Ельцина», тел. +7 343 3754 145, e-mail: OM-Zvereva2008@yandex.ru, ORCID ID: 0000-0002-3897-1465.

Шабунин Сергей Николаевич
д.т.н., профессор, зав. кафедрой радиоэлектроники и телекоммуникаций, ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б. Н. Ельцина» (УрФУ, 620002, Екатеринбург, ул. Мира, 19), e-mail: S.N.Shabunin@urfu.ru, ORCID ID: 0000-0003-2862-8812.

Авторы прочитали и одобрили окончательный вариант рукописи.
Авторы заявляют об отсутствии конфликта интересов.
Вклад соавторов: *Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.*

Technique for Building Project Teams Based on the Belbin Role Model

Ekaterina V. Vershinina, Olga M. Zvereva, Sergey N. Shabunin

Ural Federal University named after the first President of Russia B. N. Yeltsin (UrFU)

Abstract: This article presents a technique for building teams capable of successfully implementing project activities. Usually, when project teams are to be built, only the functional component is taken into account, i.e. team members must have the professional knowledge and skills necessary to create a specific product. Experience shows that this does not guarantee success. On the other hand, there are different techniques for determining the individual characteristics of people, but they are, as a rule, confined to the theory and practice of management, and are aimed at building teams of specialists with a certain profile without taking into account differences in professional competencies. A distinctive feature of the proposed methodology is the combination of considered professional competencies and skills defined as specializations and the individual psychological characteristics of potential team members. At the same time it is proposed to determine these individual characteristics on the basis of Belbin role model because this model is focused on the successful implementation of projects. The developed application using the data about candidates generates all possible teams. For each team a role diversity indicator is estimated in accordance with the Belbin model, this indicator is considered to be the assessment of the potential effectiveness of the team. As a result, the team for which this indicator has the maximum value is recommended to run the project. The presence of a participant with leadership qualities is also an important success factor, so the application checks and prints a diagnosis about whether there is a candidate with this role in the built team. An application that automates this technique provides an additional opportunity to view data about team members

and delivers the ability to edit its composition. The technique can be applied both in companies where there is a possibility of choosing from several employees with similar specialization, and also in the educational process where project team work becomes one of the main forms of students educational activities, and where different team compositions are possible as well. It was successfully tested in one of the IT companies in the Ural region, and also proved its promise in the educational teams building in the Department of Information Technologies and Automation and at the Department of Radio Electronics and Telecommunications, Institute of radioelectronics and information technologies, Ural Federal University named after the first President of Russia B.N. Yeltsin.

Keywords: team, web application, project, lifecycle, the Belbin model, educational process, software tool.

For citation: Vershinina E. V., Zvereva O. M., Shabunin S. N. Technique for building project teams based on the Belbin role model (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 3, pp. 14-27. https://doi.org/10.55648/1998-6920-2024-18-3-14-27.

Content is available under the license Creative Commons Attribution 4.0 License

© Vershinina E. V., Zvereva O. M., Shabunin S. N., 2024

The article was submitted: 17.08.2023;
revised version: 06.12.2023;
accepted for publication 22.02.2024.

References

1. Belbin R. M. *Team Roles at Work*. 2nd ed. Routledge, 2010. 162 p.
2. *Rukovodstvo k svodu znanii po upravleniyu proektom (Rukovodstvo PMBOK): Institut upravleniya proektami* [A Guide to the Project Management Body of Knowledge (PMBOK Guide): Project Management Institute]. Project Management Institute, 2017. 370 p.
3. Katzenbach J. R., Smith D. K. *The Wisdom of Teams: Creating the High-Performance Organization*. Harvard Business Review Press; Reprint edition (October 13, 2015). 304 p.
4. Krychevsky R. L. *Psihologiya liderstva* [Psychology of Leadership]. Moscow, Staut, 2007. 541p.
5. Winston W. Royce. Managing the Development of Large Software Systems. *Technical Papers of Western Electronic Show and Convention (WesCon)*, 1970, available at: http://leadinganswers.typepad.com/leading_answers/files/original_waterfall_paper_winston_royce.pdf (accessed 22.07.2023).
6. Sennik U. S., Grebennikov R. I. Ghiznenny cikl informacionnyh cictem [The life cycle of information system]. *Sistemnyj analiz i prikladnaya informatika*. 2015, no 2. pp. 4-9.
7. Volfson B. *Gibkiye metodologii razrabotki* [Life cycle of information systems], available at: https://strategium.space/wp-content/uploads/2018/07/Gibkie-metodologii.pdf (accessed 11.07.2023).
8. GOST R ISO/IEC 12207-2010. Information technology. System and software engineering. Software life cycle processes. [Information technology. Set of standards for automated systems. Automated systems. Stages of development]. *Natsional'nyi standart Rossiiskoi Federatsii*. Moscow, 2011. 100 p.
9. GOST R ISO/MEK 12207-2010. Informatsionnaya tekhnologiya. Sistemnaya i programmnaya inzheneriya. Protsessy zhiznennogo tsikla programmnykh sredstv [GOST R ISO/IEC 12207-2010. Information technology. System and software engineering. Software life cycle processes]. *Natsional'nyi standart Rossiiskoi Federatsii*, Moscow, Standartinform, 2011. 100 p.
10. Dyaram L. and Kamalanabhan T. J. Diversity in software development teams: the moderating effects of interdependence, team tenure and project type. *International Journal of Business Innovation and Research*, 2011, vol. 5, no 2, pp. 159-178.
11. Hsu B. F., Wu W. L., Ye R. S. Team personality composition, affective ties and knowledge sharing: a team-level analysis. *Int. Journal of Technology Management*, 2011, vol. 53, no 2-4, pp. 331-351.
12. He Z. and Thatchenkery T. Understanding culture-specific leadership relationship in a multi-cultural virtual project team (MVPT): a case study. *International Journal of Business and Globalisation*, 2011, vol. 7, no 4, pp. 446-458.
13. Lech M. B., Georgiade T. J., Sukup J. E., Tomovic C. L. The impact of Product Lifecycle Management on virtual team creation: a concept change model for heightened productivity. *International Journal of Manufacturing Technology and Management*, 2020, vol. 19, no 3-4, pp. 279-293.

14. Parker G. M. *Team Players and Teamwork: The New Competitive Business Strategy*. Wiley, 1990. 178 p.
15. Davis J., Millburn P., Murphy T., Woodhouse M. *Successful Team Building: How to Create Teams that Really Work*. Kogan Page, 1992. 184 p.
16. Prichard J. S. and Stanton N. A. Testing Belbin's team role theory of effective groups. *Journal of Management Development*, 1999, vol. 18, no 8, pp.652-665.
17. Park W-W. and Bang H. Team role balance and team performance. *Belbin's Biennial Conference, Changing Role of Management in the 2021st Century*, Cambridge: Clare College, 2002.
18. Zarzu C., C. Scarlat C., Falcıoğlu P. Team Composition and Team Performance: Achieving Higher Quality Results in an International Higher Education Environment. *Int. Conf. "Active Citizenship by Knowledge Management & Innovation"*, Zadar, Croatia, 19-21 June, 2013, pp. 1321-1328.
19. Zvereva O. M., Shams R. Software Support for Team Engineering: Educational Case for IT Students. *2018 IV Int. Conf. on Information Technologies in Engineering Education (Inforino)*, Moscow, 23-26 October, pp. 113-118, available at: https://nrl.northumbria.ac.uk/id/eprint/40337/1/AAM_Zvereva%26Shams_2018%20Inforino%20Conference_Software%20Support%20for%20Team%20Engineering.pdf (accessed 17.02.2024).
20. Partington D. and Harris H. Team role balance and team performance: an empirical study. *Journal of Management Development*, 1999, vol. 18, no 8, pp. 694-705.
21. Senior B. Team roles and team performance: Is there 'really' a link? *Journal of Occupational and Organizational Psychology*, 1997, vol. 70, pp. 241-258.
22. *V poiskah komandy: Belbin v kratkom izlozhenii s dopolneniem* [In search of a team: Belbin in a summary with an addendum]. HR-Academy, available at: https://hr-academy.ru/hrarticle/v-poiskah-komandyi--belbin-v-kratkom-izlojenii-s-dopolneniem.html (accessed 17.07.2023).
23. Cepkov M. Tipologiya rolej v komande [Typology of roles in the team]. *Mytype*, iss. 2, Typology. 2016, no. 1-2, available at: https://mtsepkov.org/ Типология_ролей_в_команде_(статья_в_MyType_1-2.2016) (accessed 18.07.2023).
24. Model' Belbina dlya IT: sila i slabost' raznyh komand [Belbin's model for IT: the strengths and weaknesses of different teams]. *Habr*, 2020, available at: https://habr.com/ru/company/oleg-bunin/blog/522586/ (accessed: 22.07.2023).

Ekaterina V. Vershinina
Student, Ural Federal University named after the first President of Russia B. N. Yeltsin, e-mail: Ek_Vershinina@mail.ru, ORCID ID: 0009-0008-1690-9143.

Olga M. Zvereva
PhD (Engineering), Associate Professor of the Department of Information Technologies and Automation, Ural Federal University named after the first President of Russia B. N. Yeltsin, phone: +7 343 3754 145, e-mail: OM-Zvereva2008@yandex.ru, ORCID ID: 0000-0002-3897-1465.

Sergey N. Shabunin
Dr. of Sci. (Engineering), Professor, Head of the Department of Radioelectronics and Telecommunications, Ural Federal University named after the first President of Russia B. N. Yeltsin (UrFU, Russia, 620002, Ekaterinburg, Mira St., 19), e-mail: S.N.Shabunin@urfu.ru, ORCID ID: 0000-0003-2862-8812.

DOI: 10.55648/1998-6920-2024-18-3-28-44 УДК 535.21

Анализ эффективности генерирования и преобразования оптических ФИМД-сигналов на базе классического RoF-сегмента, основанного на применении двух лазеров

И. Л. Виноградова[1], Е. Ю. Головина[2]

[1] ФГБОУ ВО «Уфимский университет науки и технологий»
[2] Институт нефтепереработки и нефтехимии ФГБОУ ВО «УГНТУ» в г. Салавате

Аннотация: В статье предлагается метод генерации и преобразования оптических ФИМД-сигналов на базе классического RoF-сегмента, основанного на применении двух лазеров. Произведена оценка параметров схемы классического RoF-сегмента, основанного на применении двух независимых лазерных источников. Предложен подход к моделированию параметров распространяющегося излучения. Установлено, что в случае отсутствия чирпирования сигналов данная схема может быть применена для задач получения радиочастоты в пределах до нескольких ГГц. С ростом частотного удаления оптических частот этих лазеров существенно растет уровень нестабильности, сводя на нет достоинства оптического генерирования радиочастот из-за возрастания фазового шума. Тем не менее определен диапазон параметров, при котором рассматриваемая схема обеспечивает относительно стабилизированный режим работы, а значит, может найти применение в задачах промышленного интернета (IIoT), внедрение которого на сегодняшний день происходит во многих компания.

Ключевые слова: Radio-over-Fiber, волоконно-оптические линии передачи, диапазон частот, лазерные диоды, модулятор Маха–Цендера, промышленный интернет, программный пакет Maple 12, ФИМД-сигналы, чирпирование, фазовые шумы

Для цитирования: Виноградова И. Л., Головина Е. Ю. Анализ эффективности генерирования и преобразования оптических ФИМД-сигналов на базе классического RoF-сегмента, основанного на применении двух лазеров // Вестник СибГУТИ. 2024. Т. 18, № 3. С. 28–44. https://doi.org/10.55648/1998-6920-2024-18-3-28-44.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Виноградова И. Л., Головина Е. Ю., 2024

Статья поступила в редакцию 14.12.2023;
переработанный вариант – 31.01.2024;
принята к публикации 06.03.2024.

1. Введение

Развитие роботизированных систем и искусственного интеллекта привело к появлению такой концепции в инфокоммуникационной отрасли, как интернет вещей (Internet of Things, IoT), включая промышленный интернет (Industrial Internet of Things, IIoT). Всё это в итоге сейчас формирует такую методологию построения систем связи «последней мили», как «интернет всего», когда взаимодействие происходит не только между удаленным оператором (человеком) и устройством (машиной, аппаратом), но и между автоматическими устройствами без участия человека. Очевидно, что это вызывает резкий рост количества устройств, подключенных к телекоммуникационной сети, с весьма разнородным трафиком [1]. На сего-

дняшний день существуют различные технологии передачи данных, такие как Wi-Fi, LTE (Long Term Evolution), Radio-over-Fiber (RoF) и др.

Беспроводная сеть Wi-Fi позволяет устройствам подключаться и обмениваться информацией на расстоянии до 100 метров. В сети Wi-Fi используется два свободных диапазона частот: 2.4 ГГц и 5 ГГц. Технология Wi-Fi предоставляет большую скорость передачи данных и обладает низкой задержкой, но имеет малый радиус действия. К тому же для подключения Wi-Fi-устройств к волоконно-оптической линии передач (ВОЛП) требуются дополнительные преобразующие устройства, при введении в эксплуатацию которых требуется инсталляция.

Технология LTE обеспечивает связь на значительные расстояния. Изначально данная технология была разработана как элемент сетей связи общего пользования для предоставления услуг подвижной связи. С появлением IoT проводились работы по модернизации стандартов LTE, но технологии этого типа по-прежнему характеризуются высоким энергопотреблением из-за сложных механизмов сигнализации и обеспечения безопасности. Большинство этих функций не требуется для приложений IoT и взаимодействия с большим количеством устройств [2].

Технология RoF является эффективным способом передачи информации с наиболее упрощенной аппаратной частью в сравнении с Wi-Fi и LTE, что обеспечивает минимум энергопотребления и миниатюризацию. Фактически в RoF имеет место аналоговый способ формирования радиосигнала как разностной частоты в результате сложения в нелинейном элементе оптических сигналов на близких несущих. Это позволяет разворачивать RoF-системы буквально «на ходу», что весьма актуально для систем связи специального назначения (военных, МЧС и т.п.). Можно сказать, что RoF совмещает достоинства волоконно-оптических систем с гибкостью и мобильностью, присущими радиосвязи. Данная технология позволяет передавать данные по оптическим волокнам на большие расстояния, а затем преобразовывать их в радиосигналы для беспроводной связи. Это обеспечивает высокую пропускную способность и низкую задержку передачи данных, что важно также для поддержки массового количества устройств в такой сети, как IoT.

Информационная технология RoF изначально основана на принципе получения радиочастотного сигнала гигагерцового диапазона частот посредством фотодетектирования сложного (суммарного) оптического сигнала, содержащего данную радиочастоту как разностную [3], рис. 1. В случае формирования сигнала путем сложения оптических волн от лазерных диодов (LD – laser diode) непрерывного излучения LD1 и LD2 с близкими центральными длинами волн λ_1 и λ_2 (круговыми частотами ω_1 и ω_2 соответственно) вследствие биения частот возникают составляющие: $\omega_1 \pm \omega_2$ [3, 4]. При условии, что ω_3 находится в рабочем диапазоне амплитудно-частотной характеристики (АЧХ) фотодиода (PD – photodetector), а ω_1 и ω_2 лежат в области его спектральной восприимчивости, после PD наблюдается разностная круговая частота: $\omega_3 = |\omega_1 - \omega_2|$.

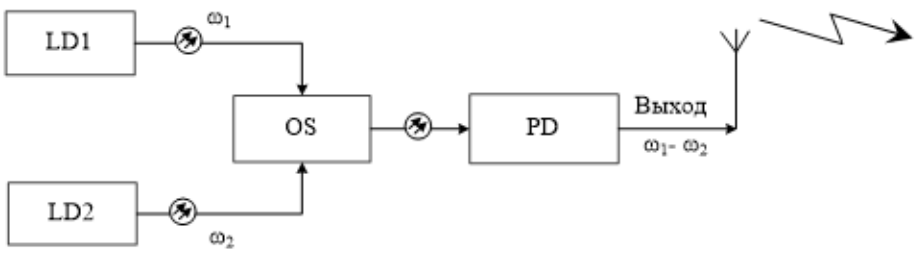


Рис. 1. Генерация разностной частоты ввиду эффекта биения двух оптических волн.
Обозначения: OS – Optical Splitter – оптический разветвитель

Именно это и оправдывает само название Radio-over-Fiber, т.е. передачу радиочастоты (в данном случае ω_3) непосредственно за счет физических эффектов, имеющих место в оптоволокне (ОВ). Если после PD подключить антенный радиоизлучатель, то ω_3 будет излучаться в радиоэфир. Если излучение одного из лазеров ω_1 является непрерывным, а излучение вто-

рого лазера ω_2 модулировано информационным сигналом с полосой $\Delta\omega$ (т.е.: $\omega_2 \pm \Delta\omega/2$ в оптическом канале), то получаемый на выходе фотодиода (а также и в радиолинии) сигнал будет иметь спектр $\omega_3 \pm \Delta\omega/2$.

Ценность таких схем состоит в аппаратной простоте метода генерации радиочастоты f (практически в любом высокочастотном (ВЧ) – радиодиапазоне) с обеспечением заданной полосы канала Δf, простоте перестройки параметров f и Δf в значительных пределах (кратное изменение в разы путем незначительного (проценты) изменения оптических параметров схемы). Поэтому при разработке RoF-схемы для генерирования и преобразования оптических сигналов с функцией изменения мгновенной длины волны (ФИМД) (с трансляцией соответствующих параметров в радиодиапазон) для приложений изначально представляется логичным рассмотреть схему с двумя независимыми лазерами (рис. 1). Эта схема обеспечивает получение радиочастоты в наиболее широком диапазоне (до терагерцовой области) и вполне может использоваться совместно с MZM (MZM – Mach–Zehnder Modulator).

Несмотря на то, что известны работы, в частности [5, 6], в которых в результате экспериментальных исследований подобной схемы делается вывод о значительном дрожании несущей радиочастоты f (возникающей вследствие разности использующихся оптических длин волн), увеличивающемся по мере роста f. Дословно[1] авторы пишут, что схемы некогерентной генерации приводят к тому, что радиочастотный сигнал несет суммарный фазовый шум обоих лазеров, и, таким образом, вероятно, фазовый шум выходит за рамки правил ITU для радионосителей [7], если только лазеры не имеют узкую ширину полосы пропускания и жестко контролируются [5]. Схемы самогетеродина, в которых свет от одного лазера разделяется на два оптических тона путем модуляции синусоидой, обеспечивают значительно лучшую фазовую стабильность [5, 6] ценой более сложной генерационной установки. Но в перечисленных работах рассматривается сложный многокомпонентный волоконно-оптический сегмент, кроме того, не получены ограничения на значения радиочастоты и схемные ограничения для ВОЛП-RoF, для которых, возможно, схема на рис. 1 может быть признана работоспособной. Еще раз подчеркнем такое достоинство последней, как аппаратная простота (что, кстати, не отрицают и авторы [5, 6]), что весьма актуально для сегментов промышленного интернета, обслуживающих каналы передачи мониторинговой информации. Поэтому представляется интересным проанализировать возможность применения упомянутой схемы (рис. 1) в применении для сегментов промышленного интернета.

2. Материалы и методы

Выполнение ФИМД (чирпирования) оптических сигналов может быть реализовано в различных её частях, рис. 2: а) в канале[2] включения первого лазера (λ_1), б) второго лазера (λ_2), в) на входе ВОЛП-RoF и г) на выходе ВОЛП-RoF. Как известно, ФИМД-сигнал может использоваться для различных целей – для компенсации дисперсии, тогда более актуальными являются схемы а) ÷ в) на рис. 2, или для формирования радиочастотных свойств канала средствами волоконной оптики, для чего важен ФИМД-сигнал на выходе ВОЛП-RoF, подаваемый непосредственно на радиоизлучающее устройство, т.е. схема г) на рис. 2. Для последней упомянутой задачи можно использовать и первые три схемы, если известны свой-

[1] Incoherent generation schemes result in the RF signal carrying the combined phase noise of both lasers – and thus likely a phase noise that is outside of ITU regulations for radio carriers [7], unless the lasers are of narrow linewidth and tightly controlled [5]. The self-heterodyne schemes where the light from a single laser is split into two optical tones by modulating it with a sinusoid, provide significantly better phase stability [5,6] at the cost of a more complex generation setup.

[2] Чирпирование может достигаться как с применением отдельного блока (устройства, схемы), основанного, например, на применении нелинейных оптических эффектов, входящего в состав сегмента, как показано на рис. 2, так и с применением чирпирующего MZM либо специализированного интерфейса управления лазером.

ства ВОЛП-RoF. Но в любом случае следует получить оценку влияния шумов лазеров, чтобы определиться с областью применения данной схемы.

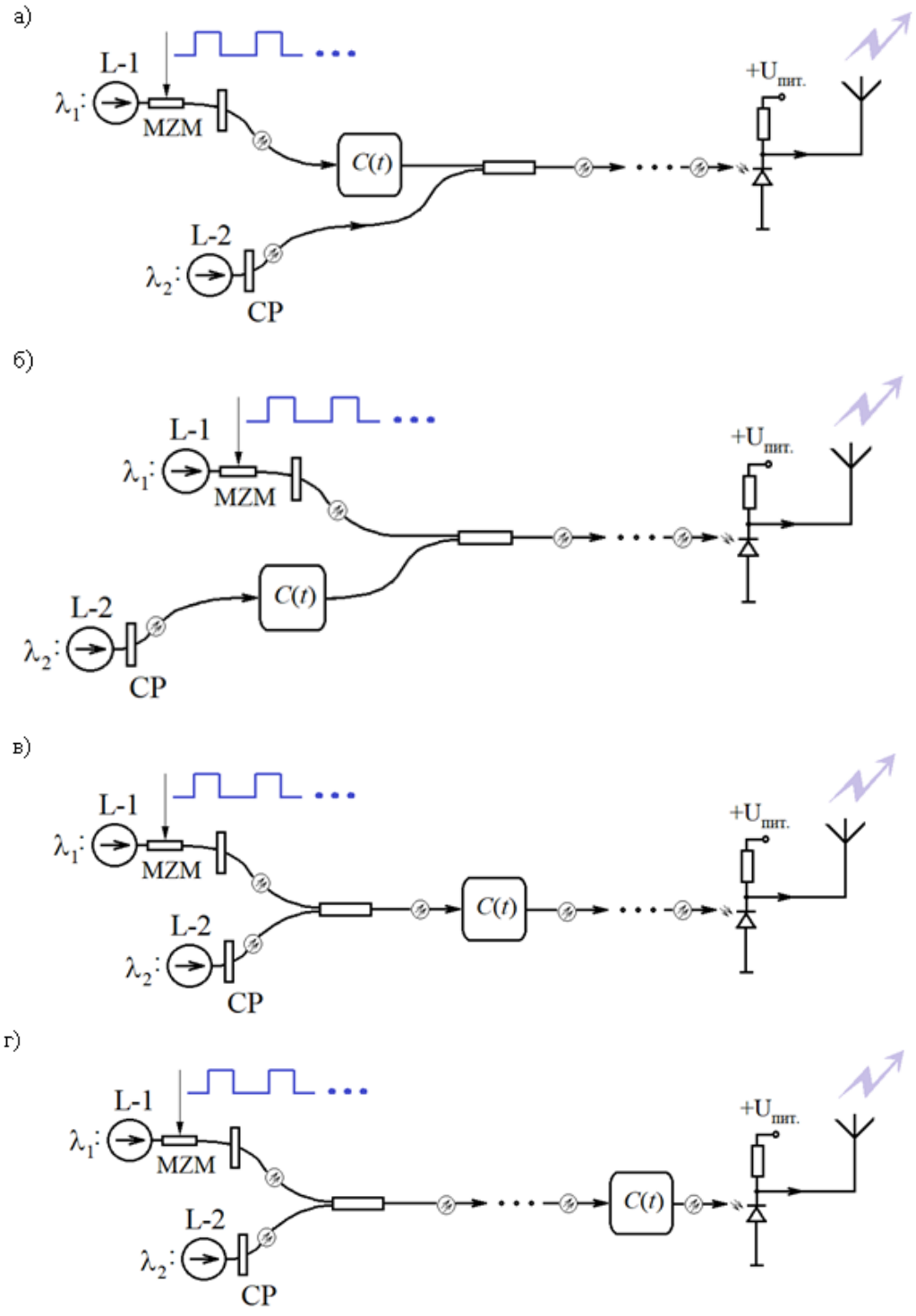


Рис. 2. Иллюстрация расположения блока (устройства, схемы), обеспечивающего получение ФИМД-сигнала в RoF-сегменте (обозначено C(t)); CP – контроллеры поляризации

Так, согласно [8, 9] основными неидеальностями (шумами, искажениями) полупроводниковых лазеров являются:

1) предел монохроматичности, или конечная ширина линии излучения;

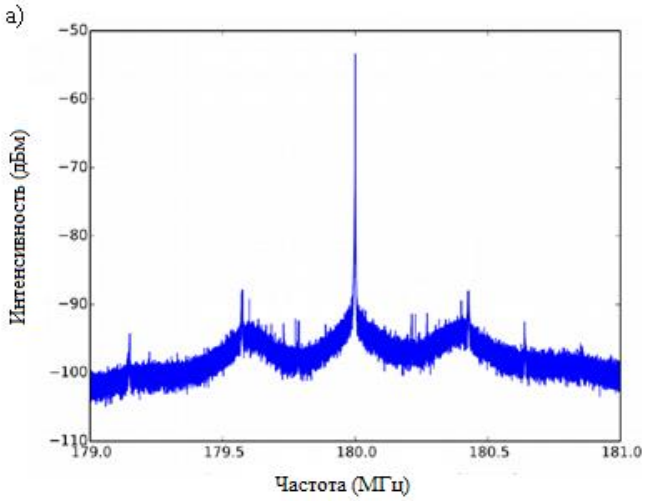
2) блуждание центральной частоты основной генерируемой моды, что связано с изменениями длины резонатора, вызванными вибрациями, флуктуациями температуры и давления, проявляющее как кратковременный ($\sim 10^{10} \div 10^{12}$ $с^{-1}$), так и долговременный ($\sim 10^{3}$ $с^{-1}$ или менее) характер;

3) фазовый шум, обусловленный тем, что каждый самопроизвольно излучаемый фотон добавляет к когерентному полю (определяемому стимулированным излучением) небольшую компоненту, фаза которой является случайной, что и возмущает фазу поля случайным образом. Кроме того, при изменении оптического усиления (при вариации населенности носителей) показатель преломления активной среды также изменяется, что сопровождается фазовой модуляцией (и в итоге также приводит к конечной ширине спектральной линии);

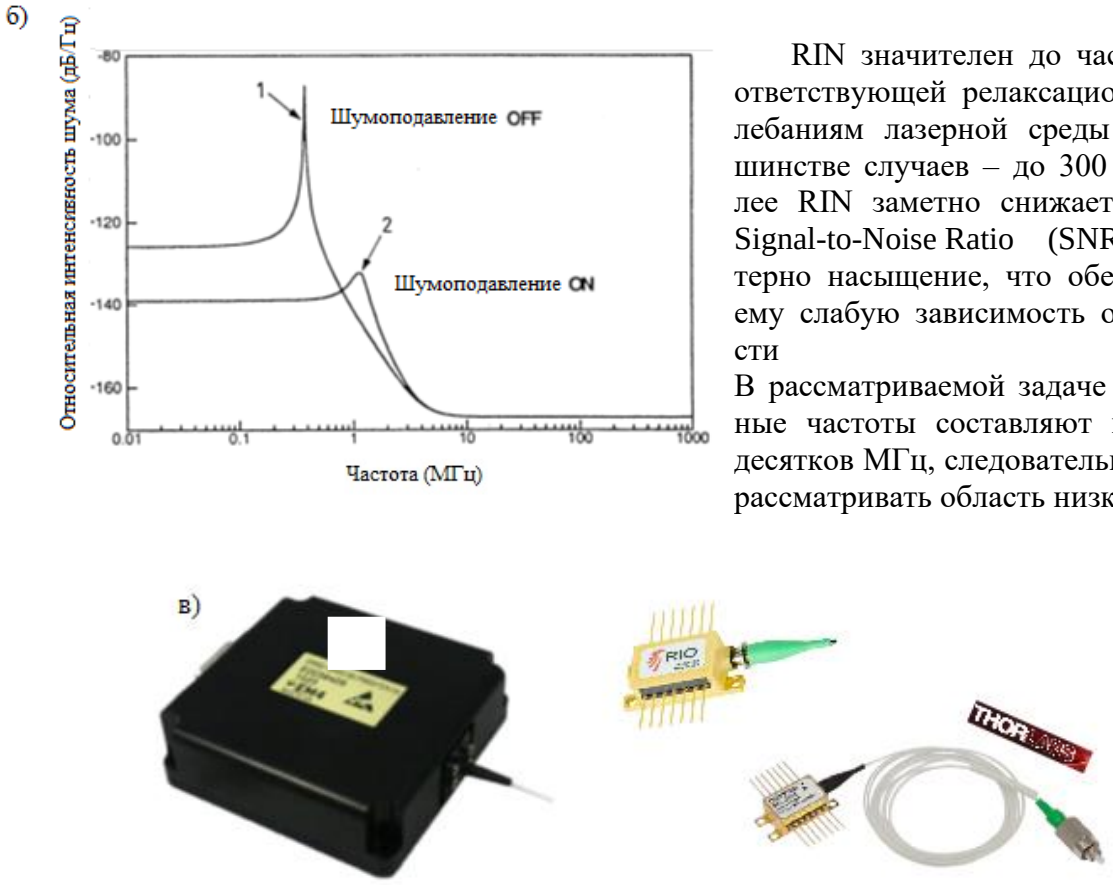
4) флуктуации тока смещения и связанное с ним блуждание амплитуды, вызванные спонтанностью характера излучения и шумом электронно-дырочной рекомбинации (дробовой шум) как кратковременный процесс, а также и долговременный дрейф выходной мощности, возникающий из-за теплового смещения резонатора, деградации зеркал и других компонентов лазера, включая изменение параметров активной среды (relative intensity noise – RIN);

5) шум модового распределения MPN (антифазная динамика [9]), связанный с тем, что при генерации нескольких мод, даже если общая мощность остается постоянной, мощность в каждой моде (основной и малых боковых[3], которые могут быть даже на 20 дБ ниже основной, рис. 3а) колеблется во времени, что представляет серьезную проблему для поддержания интенсивности в каждой моде. При этом мощность боковых мод колеблется от 0 до их наибольшего значения и вызывает соответствующее колебание мощности основной моды при практически незаметной интенсивности результирующего излучения (что значительно увеличивает RIN основной моды на частотах ниже его максимума в диапазоне 0–1 ГГц, рис. 3б. Физическое происхождение боковых мод связано с релаксационными колебаниями [9].

Если рассматривать задачу использования ФИМД-сигнала для управления фазированными антенными решетками (ФАР) на сегментах промышленного интернета, тогда стабилизация частотных свойств лазеров особенно актуальна. Однако в этом направлении в настоящее время разработаны и являются коммерчески доступными большое количество моделей лазеров с общей шириной линии (включая блуждания центральной частоты) в пределах до 1 кГц [10–23], рис. 3в, что представляется вполне достаточным. Известны даже модели с шириной линии существенно менее 1 кГц.



[3] Для значительного количества моделей DFB-лазеров характерны дополнительные низкие продольные моды, в случае VCSEL (vertical-cavity surface-emitting laser) это могут быть поперечные моды.



RIN значителен до частоты, соответствующей релаксационным колебаниям лазерной среды (в большинстве случаев – до 300 кГц). Далее RIN заметно снижается, а для Signal-to-Noise Ratio (SNR) характерно насыщение, что обеспечивает ему слабую зависимость от мощности [9]. В рассматриваемой задаче характерные частоты составляют несколько десятков МГц, следовательно, можно рассматривать область низкого RIN.

Рис. 3. Параметры полупроводниковых лазеров: (а) спектр излучения ультраузкополосного DFB-лазера [23]; (б) типичные спектры шума относительной интенсивности (RIN) для лазера с диодной накачкой без (кривая 1) и с активным шумоподавлением (2) [11] с отрицательной обратной связью; (в) модели коммерчески доступных ультраузкополосных лазеров

Так, в [23] экспериментально установлена ширина линии в пределах до 200 Гц, авторы об этом пишут следующее[4]: демонстрируется миниатюрный DFB-лазер с самоинжекционной блокировкой, использующий резонансную оптическую обратную связь от высокодобротного кристаллического резонатора в режиме шепчущей галереи. Коэффициент уменьшения ширины полосы пропускания составляет более 10.000, а мгновенная ширина полосы пропускания в результате не превышает 200 Гц. Минимальное значение отклонения Аллана для стабильности частоты лазера составляет 3×10^{-12} при времени интегрирования 20 мкс. Лазер обладает превосходной спектральной чистотой и хорошей долговременной стабильностью.

Безусловно, для таких лазеров требуется термостабилизация лучше, чем 0.01°C [10 и др.], а также желательна виброизоляция. Указанные характеристики достигаются такими факторами, как: введение обратной связи с частотным дискриминатором, использующим атомный или молекулярный резонанс[5], оптогальванический эффект[6], применение контура фазовой синхронизации с интерференционной подстройкой[7], фильтров с гребенчатым пери-

[4] We demonstrate a miniature self-injection locked DFB laser using resonant optical feedback from a high-Q crystalline whispering gallery mode resonator. The linewidth reduction factor is greater than 10.000, with resultant instantaneous linewidth less than 200 Hz. The minimal value of the Allan deviation for the laser frequency stability is 3×10^{-12} at the integration time of 20 μs. The laser possesses excellent spectral purity and good long term stability.

[5] Для привязки частоты выходного излучения к резонансной частоте лазера.

[6] Для фиксации частоты лазера до атомного или молекулярного резонанса, что имеет место в MQW-лазерах, особенно для напряженных квантовых ям [18]. Это же справедливо и для предыдущего случая.

[7] С интерферометром Майкельсона, откалиброванным посредством стабилизированного мастер-лазера DFB, что обеспечивает набор одинаково расположенных опорных частот, либо эталона Фабри-Перо (FP) с контуром контроля температуры эталона.

одическим спектром (например, AWG) с эталонным набором частот. Как правило, это лазеры FP-DFB или MQW-DFB.

Для того чтобы оценить влияние фазового шума лазеров для схемы на рис. 2, определим ток фотодиода J_{PD} (а значит, и мощность радиоизлучения, учитываем также, что фотодиод, как правило, обладает квадратичной передаточной функцией) через амплитуды[8] излучения лазеров:

$$J_{PD} = K_{PD} \cdot (E_1 + E_2)^2, \quad (1)$$

где $E_{1,2} = E_{01(02)} \cdot \left(\cos\left(\omega_{1,2}t + \left\langle\varphi_{1,2}(t)\right\rangle\right) + \sin\left(\omega_{1,2}t + \left\langle\varphi_{1,2}(t)\right\rangle\right)\right)$, K_{PD} – коэффициент оптоэлектронного преобразования фотодиода $\left\langle\phi_{1,2}(t)\right\rangle$ – усредненные величины фазовых отклонений.

Если раскрыть квадрат суммы, выполнить тригонометрические преобразования и учесть при этом, что $(E_1)^2$ и $(E_2)^2$ определяют интенсивности соответствующих излучений, получим:

$$J_{PD} = K_{PD} \cdot \left(I_1 + I_2 + 2E_1E_2 * \cos(|\omega_1 - \omega_2|) * \left(1 - \sin\left(\left|\left\langle\varphi_1(t)\right\rangle + \left\langle\varphi_2(t)\right\rangle\right|\right)\right)\right). \quad (2)$$

При выводе (2) учтено соотношение для косинуса разности, а также приравнены нулю деструктивные (не имеющие физического смысла) члены вида: $2E_{01(02)}\cos\left(\omega_{1,2}t\right)\cdot\sin\left(\omega_{2,1}t\right)$. В (2) также предполагается, что круговые частоты ω_1 и ω_2 являются близкими, соответствуют λ_1 и λ_2, и обе длины волны приблизительно одинаково воспринимаются фотодиодом (K_{PD1} = =K_{PD2} = K_{PD}). Значения фазовых отклонений $\left\langle\phi_{1,2}(t)\right\rangle$ несложно найти из технических данных лазеров: $\left\langle\phi_{1,2}(t)\right\rangle \cong \varphi\cdot\Delta\nu$, где φ представляет параметр PhN и равен ≅20 μrad/Hz для рассматриваемых типов моделей лазеров [21, 22]. Если Δν = 1 кГц, то можно говорить о фазовых скачках порядка 10^{-2} рад, которые для времен порядка 10^{-12} с (субтерабитные скорости передачи информации; учитываем также, что ω имеют порядок 10^{14} рад/с) приведут к заметному (т.е. того же порядка) блужданию радиочастоты $\Omega = \omega_1 - \omega_2$. Но на сегментах промышленного интернета в настоящее время такие битовые скорости маловероятны, следовательно, фазовый шум лазеров скажется существенно (на 1–2 порядка) меньше, и им в первом приближении можно пренебречь.

При выполнении анализа влияния амплитудного шума на работоспособность схемы на рис. 2 следует учитывать также и возможный отраженный сигнал на ВОЛП-RoF, хотя большинство современных моделей лазеров снабжены оптическими изоляторами. Последние значительно снижают попадание отражений в лазер, но не компенсируют его полностью (остаточный уровень находится в пределах ≈ -70 дБ). Такое значение мало сказывалось бы для классических ВОЛП, но в рассматриваемой схеме лазеры оказываются связанными по интенсивности. RIN также вносит сюда свой вклад.

Так, в [24, 25] показано, что для инженерного моделирования интенсивности оптического сигнала на выходе двух связанных (по интенсивности) лазеров допустимо использовать следующие уравнения:

$$\frac{\partial I_1(t)}{\partial t} = \gamma P_1(t) G_n - I_1(t) - k_{11} I_1^2(t) - k_{21}\bar{I}_1(t-\tau) I_2(t),$$
$$\frac{\partial I_2(t)}{\partial t} = \gamma P_2(t) G_n - I_2(t) - k_{22} I_2^2(t) - k_{12}\bar{I}_2(t-\tau) I_1(t), \quad (3)$$

[8] Речь идет о напряженностях электрического поля световых волн E_1 и E_2, вызывающих фотоэффект. В (1) не учитываются потери на линейное затухание и дисперсия ввиду малой длины ВОЛП-RoF.

где k_{lm} – коэффициенты инжекционной взаимосвязи между процессами в резонаторах лазеров; P_1 *и* P_2 – параметры накачки: $P_{1,2}(t) = \frac{J_{1,2}(t)}{J_{th}} - 1$, где $J_1(t)$ и $J_2(t)$ – токи накачки лазеров соответственно, и J_{th} – пороговый ток накачки для лазера без обратной связи (считаем эти значения одинаковыми для рассматриваемых лазеров); G_n – характеризует динамическое модовое усиление (считаем, что оба лазера похожи конструктивно): $G_n = \frac{G_0}{1 + \delta\omega^2 \cdot {\tau_s}^2 + \frac{I(t)}{I_{0,n}}}$.

Здесь $\delta\omega$ – разность частот поступающего оптического излучения и рабочего квантового перехода электрона; τ_s – время жизни неравновесных носителей; $I_{0,n}$ – интенсивность насыщения (также считаем, что оба лазера обладают примерно одинаковыми динамическими свойствами). Текущее время нормировано на время жизни фотонов в резонаторе: t / τ_{th}, T – отношение времен жизни неравновесных носителей и фотонов, т.е. $T = \frac{\tau_s}{\tau_{th}}$. В (3) учтено, что изменение интенсивности прямо пропорционально (γ) параметрам накачки и параметрам усиления. Будем считать, что ВОЛП-RoF не является длинной, следовательно, дисперсией и погонным затуханием можно пренебречь.

3. Результаты

Численное моделирование динамического характера изменения интенсивностей излучения для различных значений отстройки $\delta\omega = |\omega_1 - \omega_2|$ на выходе короткой ВОЛП-RoF проведено для DFB-лазеров с параметрами: τ_s =1 нс, τ_{th} = 1 пс, J = 2.2 J_{th}. При выборе значений τ_s и τ_{th} предполагалось, что длительность τ_{th} определяется эффективным количеством переотражений в резонаторе Фабри–Перо DFB-лазера, умноженным на длительность одного прохода; длительность τ_s приблизительно определяется длительностью импульса в системе передачи, а ток накачки J должен обеспечивать работу лазера в когерентном режиме, т.е. кратно превосходить пороговый ток J_{th}. Значение, равное 2.2, принято для моделирования как эффективно усредненная величина.

Для получения результатов моделирования было применено численное интегрирование уравнений (3) с использованием программного пакета Maple 12. На рис. 4 представлены графические зависимости относительных интенсивностей (отнесенных к усредненному за интервал наблюдения значению $\overline{I_M}$) как функций времени (а), а также характера изменения величины разности $\Delta I = |I_1 - I_2|_{\max}$, тоже[9] отнесенного к $\overline{I_M}$, и характера изменения со временем периода осцилляций Ξ для функций $I_1(t)$ и $I_2(t)$ – рис. 4б. Отличие кривых на рис. 4а от строго гармонических связано, по-видимому, с нелинейным характером уравнений (3) и, возможно, с погрешностями численного интегрирования. Кривые построены для различных значений ρ уровня отраженного сигнала[10] и $\delta\omega$, характеризующей различие λ_1 и λ_2. Так, на рис. 4а кривые 1 построены для ρ = 0.01, а кривые 2 – ρ = 0.02. При этом $\delta\omega$ = 6.28·10^9 рад/с, что соответствует радиочастоте в 1 ГГц. На рис. 4б все кривые построены для ρ = 0.01, но

[9] $\overline{I_{1,2,M}}$ – средние значения интенсивностей за интервал наблюдения; считаем, что для первого и второго лазера они одинаковы.

[10] Параметр ρ определяет коэффициенты k_{ij} в (3) с учетом RIN обоих лазеров.

для: 1 – $\delta\omega = 4.84\cdot10^{11}$ рад/с (77 ГГц), 2 – $\delta\omega = 6.28\cdot10^{10}$ рад/с (10 ГГц), 3 – $\delta\omega = 6.28\cdot10^{9}$ рад/с (1 ГГц) и 4 – $\delta\omega = 1.26\cdot10^{9}$ рад/с (200 МГц).

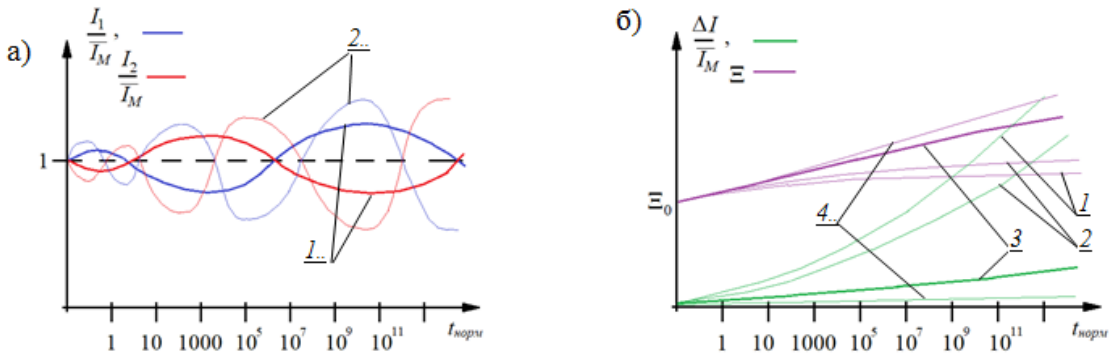


Рис. 4. Иллюстрация характера изменения интенсивностей излучения:
(а) их разностей и периода осцилляций; (б) для RoF-сегмента с двумя лазерами

Отсюда видно, что для больших значений ρ (свыше 0.01) и $\delta\omega$ (уже начиная с $6.28\cdot10^9$рад/с) система ведет себя как неустойчивая – амплитуда осцилляций увеличивается, а период осцилляций сначала немного увеличивается, а затем остается в пределах некоторой практически установившейся величины. Расчеты проведены для случая отсутствия ограничений по интенсивности излучения. Очевидно, что в реальной ситуации увеличение интенсивности ограничится её наибольшим значением, определяющимся электропитанием лазера. К тому же в расчетах не было учтено погонное затухание линии, учет которого приведет к увеличению характерного значения ρ, вызывающего колебания интенсивностей.

Рост периода осцилляций наблюдается от линейно нарастающего до асимптотически устанавливающегося. Увеличение периода (в том числе линейное нарастание) свидетельствует о переходе осцилляций к установившемуся неколебательному состоянию (т.е. к отсутствию осцилляций), а значит, имеет место стабилизация системы. Если же период со временем не увеличивается, то осцилляции в системе остаются, следовательно, система является нестабильной.

Иными словами, для RoF-сегмента с двумя лазерами, между которыми не предусмотрена конструктивная взаимосвязь для стабилизации их работы, даже без использования чирпирующих компонентов наблюдается блуждание интенсивностей излучения, приводящее и к изменению тока фотодиода (по (2), (3)), а значит, и мощности радиосигнала. В случае присутствия в линии (в том числе в радиолинии) элемента с характеристикой вида «амплитуда↔частота», что свойственно для ряда нелинейных схем, то действительно в сегменте будут присутствовать фазовые шумы, мешающие радиосвязи, что хорошо согласуется с [5, 6]. Однако эти шумы остаются весьма незначительными для ρ не выше 0.01, а также для близко расположенных λ_1 и λ_2: в частотном измерении не далее, чем в пределах нескольких сот МГц[11]. В этом случае осцилляции интенсивностей излучения со временем прекращаются (их период неограниченно возрастает). При увеличении параметра $\delta\omega$ все выглядит наоборот: система от почти стабильной переходит к нестабильной.

Чтобы оценить применимость данной схемы для рассматриваемой задачи, проведем исследование происходящих в ней процессов, но с учетом оптических ФИМД-сигналов. Последние могут быть получены: изменением тока накачки одного из лазеров, в результате включения чирпирующего MZM [26 и др.] либо устройства управления сигналами, например, такого как [27] по схемам, представленным на рис. 2. Для этого воспользуемся подходом к моделированию, аналогичным (3), но уже по отношению к напряженности поля E.

[11] Расчеты проведены для случая отсутствия блуждания центральной частоты излучения лазеров.

Так, в работе [24] Лэнг и Кобаяши ввели простую модель, описывающую работу лазера с обратной связью системой из двух дифференциальных уравнений, которые с учетом нормировки и введенного по аналогии с [28] параметра Φ, характеризующего отклонение частоты излучения для ФИМД-сигнала, приобретают вид:

$$\begin{cases} \dfrac{\partial E}{\partial s} = (1 + j \cdot a) \cdot N \cdot E + \eta \cdot e^{-j\Phi} E(s-1), \\ \dfrac{T}{\tau}\dfrac{\partial N}{\partial s} = p - N - (1 + \dfrac{2}{\tau} N)\left|E\right|^2. \end{cases} \tag{4}$$

В (4) строго согласно [24] используются следующие обозначения: s – время прохода излучения по внешнему контуру (линии), представленное (нормированное) в единицах времени жизни неравновесных носителей τ_s; a – коэффициент, связывающий текущие амплитуду и фазу поля; N – концентрация неравновесных носителей; η – коэффициент отражения; τ – время распространения излучения в резонаторе, нормированное на время τ_s; T – отношение времен жизни носителей и фотонов; p – нормированный параметр, характеризующий эффективность накачки.

Поэтому подобно тому, как это показано в [24], запишем систему уравнений для рассматриваемых лазеров с учетом Φ:

$$\begin{cases} \dfrac{\partial E_1(s)}{\partial s} = \left(1 + j \cdot \dfrac{\psi(s) \cdot E_{0,1}}{\left|E_1(s)\right|}\right) N_1(s) E_1(s) + k_{E,11} e^{-j\Delta\Phi_1(s-1)} E_1(s-1) + \\ \qquad + k_{E,21} e^{-j\Delta\Phi_2(s-1)} E_2(s-1) \left|E_1(s-1)\right|, \\ \dfrac{T}{\tau}\dfrac{\partial N_1(s)}{\partial s} = p - N_1(s) - \left(1 + \dfrac{2}{\tau} N_1(s)\right)\left|E_1(s)\right|^2; \\ \dfrac{\partial E_2(s)}{\partial s} = \left(1 + j \cdot \dfrac{\psi(s) \cdot \left|E_{0,1} - E_{0,2}\right|}{\left|E_1(s) - E_2(s)\right|}\right) N_2(s) E_2(s) + k_{E,22} e^{-j\Delta\Phi_2(s-1)} E_2(s-1) + \\ \qquad + k_{E,12} e^{-j\Delta\Phi_1(s-1)} E_1(s-1) \left|E_2(s-1)\right|, \\ \dfrac{T}{\tau}\dfrac{\partial N_2(s)}{\partial s} = p - N_2(s) - \left(1 + \dfrac{2}{\tau} N_2(s)\right)\left|E_2(s)\right|^2. \end{cases} \tag{5}$$

В (5) дополнительно приняты обозначения: $\psi(s)$ – фаза модулированного[12] излучения (переносящего информационный сигнал); $\Delta\Phi_1(s) = \dfrac{1}{\omega_1} \cdot \dfrac{\partial \psi(s)}{\partial s} - 1$; $\Delta\Phi_2(s) = \dfrac{1}{\omega_2} \cdot \dfrac{\partial \psi(s)}{\partial s} - \dfrac{\omega_1}{\omega_2}$; величины $E_{0,1}$ и $E_{0,2}$ – амплитудные значения функций $E_1(s)$ и $E_2(s)$; коэффициенты $k_{E,11}$, $k_{E,12}$, $k_{E,21}$ и $k_{E,22}$, аналогичны введенным в (3) с тем отличием, что относятся не к интенсивности, а к напряженности поля. Так же, как и выше, считаем, что динамические и мощностные характеристики лазеров приблизительно одинаковы.

[12] Предполагается, что модуляцию, т.е. введение информационного сигнала, осуществляет модулятор Маха-Цендера (MZM), располагающийся, например, в плече первого лазера, см. рис. 2.

Согласно введенной[13] и используемой в настоящей статье ФИМД-функции $C_{\omega 1}$ заметим, что $\Delta\Phi_1(s) = \dfrac{\dfrac{\partial\psi(s)}{\partial s} - \omega_1}{\omega_1} = \dfrac{c_{\omega_1}(s)}{\omega_1}$, $\Delta\Phi_2(s) = \dfrac{1}{\omega_2}\cdot\dfrac{\partial\psi(s)}{\partial s} - \dfrac{\omega_1}{\omega_2} = \dfrac{c_{\omega_1}(s)}{\omega_2}$.

При большой задержке, когда $\tau >> 1$ (т.е. когда ВОЛП-RoF много длиннее, чем резонатор лазера), можно полагать, что $\partial N/\partial s = 0$ (по аналогии с [24]), и система сводится к двум уравнениям, аналогичным (3) для каждого из рассматриваемых лазеров. Кроме того, для каждого из лазеров согласно [24] справедливо:

$$\frac{\partial E}{\partial s} = (1 + i\cdot\alpha)\cdot(p - |E|^2)\cdot E + \eta\cdot e^{-i\Phi}E(s-1). \tag{6}$$

Если по аналогии с (6) представить полученные выше уравнения (5) и учесть соотношения для функции $C_{\omega 1}(s)$, тогда запишем:

$$\begin{cases} \dfrac{\partial E_1(s)}{\partial s} = \left(1 + j*\dfrac{\psi(s)*E_{0,1}}{|E_1(s)|}\right)(p - |E_1(s)|^2 E_1(s) + k_{E,11}e^{-j\frac{c_{\omega_1}(s-1)}{\omega_1}}\cdot E_1(s-1) + \\ \qquad + k_{E,21}e^{-j\frac{c_{\omega_1}(s-1)}{\omega_2}}\cdot E_2(s-1)|E_1(s-1)|, \\ \dfrac{\partial E_2(s)}{\partial s} = \left(1 + j*\dfrac{\psi(s)*|E_{0,1} - E_{0,2}|}{|E_1(s) - E_2(s)|}\right)(p - |E_2(s)|^2 E_2(s) + k_{E,22}e^{-j\frac{c_{\omega_1}(s-1)}{\omega_2}}\cdot E_1(s-1) + \\ \qquad + k_{E,12}e^{-j\frac{c_{\omega_1}(s-1)}{\omega_1}}\cdot E_1(s-1)|E_2(s-1)|. \end{cases} \tag{7}$$

Таким образом, система уравнений (7) характеризует изменение параметров излучения в RoF-линии в случае оптического ФИМД-сигнала.

Учитывая, что интенсивность излучения представляет собой величину $I_{1,2} = \dfrac{\dot{E}_{1,2}\cdot\dot{E}_{1,2}^{*}}{2Z_B}$, где $\dot{E}^*$ – комплексно-сопряженное к $\dot{E}$, а Z_B – волновое сопротивление линии, получаем результаты моделирования, аналогичные представленным выше, см. рис. 5. Расчет проведен для ρ = 0.01, отличия частот лазерного излучения на $6.28\cdot10^9$ рад/с (1 ГГц), для следующих наибольших значений функции: $C_{\omega 1} = 6.28\cdot10^4$ рад/с (10 кГц) – кривая 1 на рис. 5а и кривая 1 на рис. 5б; $C_{\omega 1} = 1.57\cdot10^5$ рад/с (25 кГц) – кривая 2 на рис. 5б; $C_{\omega 1} = 3.14\cdot10^5$ рад/с (50 кГц) – кривая 3 на рис. 5б; $C_{\omega 1} = 4.71\cdot10^5$ рад/с (75 кГц) – кривая 2 на рис. 5а и кривая 4 на рис. 5б. В случае, если увеличить частотное расстояние между лазерами, а также и наибольшее значение ФИМД-функции, получаем существенные осцилляции интенсивностей лазеров, а также и частоту осцилляций, начинающую попадать (паразитно) в частотные диапазоны ВЧ/ШП/СШП-сигналов. Кроме того, если для малых $\delta\omega$ на рис. 4 имела место определенная стабилизация (точнее – квазистабилизация) указанных осцилляций (неограниченный рост периода Ξ, рис. 4б, например, кривая 4, говорит о снижении эффективности осцилляций), то

[13] Явление чирпирования излучения весьма по-разному представлено в литературе – начиная от неизменного параметра (коэффициента) [29], включая линейно меняющуюся величину (ЛЧМ – линейная частотная модуляция [31]), что связано с традиционно сложившимся дисперсионным характером его приобретения, и заканчивая произвольной функциональной зависимостью [30–34]. Здесь мы придерживаемся того, что ФИМД-функция может быть произвольной.

в рассматриваемом случае подобный режим удается получить только для $C_{\omega 1} \le 6.28 \cdot 10^3$ рад/с (1 кГц), что весьма мало для многих технических приложений.

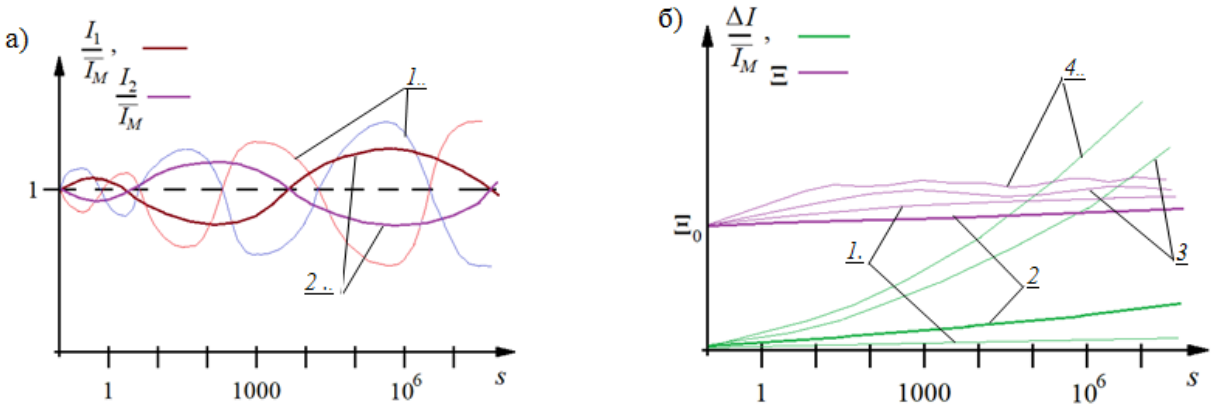


Рис. 5. Иллюстрация характера изменения интенсивностей излучения (а), их разностей и периода осцилляций (б) для RoF-сегмента с двумя лазерами, построенного по схеме рис. 2а, для случая чирпирующего MZM

Напротив, для рассмотренных значений ФИМД наблюдается характерное блуждание параметра Ξ (попеременное уменьшение и увеличение), что фактически действительно говорит о наличии фазового шума в системе. Определенная более-менее приемлемая ситуация имеет место для $C_{\omega 1} = 6.28 \cdot 10^4$ рад/с (10 кГц), что может найти техническое приложение, например, в задаче обеспечения информационной безопасности радиоканала в виде «прыгающей частоты» для нешироких канальных интервалов (до 500 МГц). Для ФИМД же со значениями $C_{\omega 1} = 4.71 \cdot 10^5$ рад/с (75 кГц) период Ξ начинает заметно блуждать и проявляется заметный рост разности интенсивностей ΔI, что фактически свидетельствует об оптическом хаосе [35] и приводит канал уже в неработоспособное состояние.

4. Обсуждение и заключение

В результате выполненного моделирования работы RoF-схемы, основанной на взаимодействии излучений двух конструктивно не связанных лазеров, можно сделать следующий вывод. Действительно, данной схеме присущи значительные фазовые шумы (что согласуется с результатами экспериментальных исследований) в случае, когда отличие центральных частот лазеров соответствует радиочастоте уже начиная от единиц ГГц и выше. Следует отметить, что указанное моделирование выполнено без учета многих дополнительно влияющих факторов, что на практике, скорее всего, еще снизит указанный порог. Тем не менее определен диапазон[14] параметров (по отличию частот лазеров и по наибольшему допустимому значению ФИМД), когда рассматриваемая схема обеспечивает относительно стабилизированный режим работы, а значит, может найти применение если не в задачах абонентских сетей (из-за малого значения радиочастоты, что на сегодняшний день неактуально), то в задачах промышленного интернета.

[14] Справедливости ради нельзя не отметить, что полученные параметры стабильной работы схемы по рис. 2 могут быть присущи либо чрезвычайно узкополосным (которые должны быть высокостабилизированными) двум лазерам, либо находятся внутри единственной полосы излучения типового (недорого) полупроводникового лазера.

Литература

1. *Виноградова И. Л., Головина Е. Ю.* Использование технологии Radio-over-Fiber в системах связи // Сборник научных статей по материалам XIII Международной научно-практической конференции «Актуальные проблемы науки и техники», Уфа, 24 ноября, 2023. С. 84–88.
2. Фам Ван Дай. Разработка моделей и методов маршрутизации в энергоэффективных ячеистых сетях дальнего радиуса действия: автореф. дис. …канд. техн. наук. СПб. 2021. 160 с.
3. *Solomon Tesfay Abraha.* Impulse Radio Ultra Wideband over Fiber Techniques for Broadband In-Building Network Application. Abstract Ph. D. thesis. Netherlands. 2012. 192 p.
4. *Haymen Shams.* Radio over Fiber Distribution Systems for Ultra Wideband and Millimetre wave Applications. Abstract Ph. D. thesis. Dublin. 2011. 158 p.
5. *Kanno A., Dat P. T., Kuri T., Hosako I., Kawanishi T., Yoshida Y., and Kitayama K.* Evaluation of frequency fluctuation in fiber wireless link with direct IQ down-converter // Proc. European Conference on Optical Communication (ECOC), 2014. P. 1–3.
6. *Gliese U., Norskov S., and Nielsen T. N.* Chromatic dispersion in fiber-optic microwave and millimeter-wave links // IEEE Transactions on microwave theory and techniques. 1996. V. 44, № 10. P. 1716–1724.
7. Recommendation ITU-R SM.1045-1. Frequency tolerance of transmitters. Electronic Publication, 1997. P. 5.
8. *Svelto O.* Principles of Lasers. Springer Science & Business Media, 2010. 620 p.
9. *Agrawal G. P.* Fiber-optic communication systems with CD. Published by John Wiley & Sons, Inc., Hoboken, New Jersey, 2010. 4nd ed. 588 p.
10. *Uekara N. and Ueda K.* 193-mHz Beat Linewidth of Frequency Stabilized Laser-Diode-Pumped Nd:YAG Ring Lasers // Opt. Lett. 18. 1993. P. 505–507. DOI: 10.1364/ol.18.000505.
11. *Scheps R.* Introduction to Diode-Pumped Solid-State Lasers. Lightwave Electronics Corp. Techn. Information, 2003. 120 p.
12. *Ikegami T., Sudo S., and Sakai Y.* Frequency Stabilization of Semiconductor Laser Diodes. Artec House. Boston, 1995. 356 p.
13. *Guy M., Villeneuve B., Latrasse C., and Tètu M.* Simultaneous absolute frequency control of laser transmitters in both 1.3 and 1.55 mu-m bands for multiwavelength communication-systems // Journal of lightwave technology. 1996. P. 1136–1143.
14. *Lee H., Lyu G. Y., Park S. Y., and Lee J. H.* Multichannel wavelength locking using transmission peaks of an AWG for multichannel optical transmission systems // IEEE Photonics Technology Letters. 1998. V.10, № 2. P. 276–278.
15. *Park Y., Lee S. T., and Chae C. J.* A novel wavelength stabilization scheme using a fiber grating for WDM transmission // IEEE Photon. Technology Letters. 1998. V. 10, № 10. P. 1446–1448.
16. *Ono T. and Yano Y.* Key technologies for terabit/second WDM systems with high spectral efficiency of over 1 bit/s/Hz // IEEE J. Quantum Electron. 1998. V. 34. P. 2080–2088.
17. *Nasu H., Takagi T., Shinagawa T., Oike M., Nomura T., and A. Kasukawa.* A Highly Stable and Reliable Wavelength Monitor Integrated Laser Module Design // Journal Lightwave Technology. 2004. V. 22. P. 1344–1351.
18. *Nasu H., Mukaihara T., Takagi T., Oike M., Nomura T., and Kasukawa A.* Wavelength monitor integrated laser modules for 25-GHz-spacing tunable applications // IEEE Journal of Selected Topics in Quantum Electronics. 2005. V.11, № 1. P. 157–164.
19. *Tissot Y., Limberger H. G., and Salathe R.-P.* Ultrawide bandwidth wavelength monitor based on a pair of tilted fiber Bragg gratings // IEEE Photon. Technology. Letters. 2007. V. 19. P. 1702–1704.
20. *Liang W., Ilchenko V. S., Savchenkov A. A., Matsko A. B., Seidel D., and Maleki L.* Whispering gallery mode resonator based ultra-narrow linewidth external cavity semiconductor laser // Optics letters. 2010. V. 35. P. 2822–2824.

21. Thorlabs. Models: EM650 High Power DFB Laser Module; SFL1550S. URL: https://www.thorlabs.com/ (дата обращения: 11.09.2023).
22. Model RIO PLANEXSeries 1550nm Low Phase Noise Narrow Linewidth External Cavity Laser, Redfern Integrated Optics (RIO). URL: http://rio-lasers.com/ (дата обращения: 01.10.2023).
23. *Consolino L., Nafa M., Cappelli F., Garrasi K., Mezzapesa F. P. et al.* Fully phase-stabilized quantum cascade laser frequency comb // Electronic and Electrical Engineering. 2018. V. 2, № 5. P. 231–237.
24. *Lang R., Kobayashi K.* External optical feedback effects on semiconductor injection laser properties // EEE J. Quantum Electron. 1980. P. 347–355.
25. *Javaloyes J., Mandel P., Pieroux D.* Dynamical properties of lasers coupled face to face // Physical review. Statistical, nonlinear, and soft matter physics. 2003. P. 11.
26. *Burdin V. A., Dashkov M. V., Volkov K. A.* Application of dispersion managed soliton regime in radio-over-fiber systems // Proceedings of SPIE. 2012. V. 8787. P. 1–6.
27. *Abdrakhmanova G. I., Andrianova A. V., Vinogradova I. L., Yantilina L. Z., Grahova E. P. et al.* Interference fiber optic device for RoF antenna radiators control // Proc. International Siberian Conference on Control and Communications, 2016. P. 1–6.
28. *Agrawal G. P.* Nonlinear fiber optics. Boston: Academic Press, 2009. 466 p.
29. *Chorchos L., Rommel S., Turkiewicz J. P., Monroy I. T., and Vegas Olmos J. J.* Reconfigurable Radio Access Unit for DWDM to W-Band Wireless Conversion // IEEE Photonics technology letters. URL: http://www.ieee.org/publications_standards/publications/rights/index.html/ (дата обращения: 08.10.2023).
30. *Meijerink A., Roeloffzen C. G. H., Meijerink R., Zhuang L., Marpaung D. A. I. et al*. Novel Ring Resonator-Based Integrated Photonic Beamformer for Broadband Phased Array Receive Antennas – Part I: Design and Performance Analysis // J. lightwave technology. January 2010. V. 28, № 1. P. 3–18.
31. *Zhuang L., Roeloffzen C. G. H., Meijerink A., Burla M., Marpaung D. A. I. et al*. Novel Ring Resonator-Based Integrated Photonic Beamformer for Broadband Phased Array Receive Antennas – Part II: Experimental Prototype // Journal of lightwave technology. January 2010. V. 28, № 1. P. 19–31.
32. *Alphones A. and Thai P. Q.* Hybrid Approach for Optical Beamforming for Phased Array // Proc. Asia-Pacific Microwave Conference, 2010. P. 311–317.
33. *Drummond M. V., Monteiro P. P., and Nogueira R. N.* Photonic True-Time Delay Beamforming Based on Polarization-Domain Interferometers // J. Lightwave technology. 2010. V. 28, № 17. P. 2492–2498.
34. *Burla M., Khan M. R. H., Marpaung D. A. I., Roeloffzen C. G. H., Maat P. et al*. Squint-Free Beamsteering Demonstration using a Photonic Integrated Beamformer based on Optical Ring Resonators // IEEE International Topical Meeting on Microwave Photonics. 2010. P. 401–404.
35. *Кившарь Ю. С., Агравал Г. П.* Оптические солитоны. От волоконных световодов до фотонных кристаллов / пер. с англ. под ред. Н. Н. Розанова. М.: Физматлит, 2005. 648 с.

Виноградова Ирина Леонидовна
д.т.н., профессор кафедры телекоммуникационных систем, ФГБОУ ВО «Уфимский университет науки и технологий» (УУНиТ, 450076, Республика Башкортостан, Уфа, ул. Заки Валиди, 32), e-mail: vil-4@mail.ru, ORCID ID: 0000-0003-4618-6843.

Головина Евгения Юрьевна
старший преподаватель кафедры информационных технологий, Институт нефтепереработки и нефтехимии ФГБОУ ВО «УГНТУ» в г. Салавате (Институт нефтепереработки и нефтехимии ФГБОУ ВО «УГНТУ» в г. Салавате, 453250, Республика Башкортостан, Салават, ул. Губкина, 22Б), e-mail: egolovina82@mail.ru, ORCID ID: 0009-0006-3624-5490.

Авторы прочитали и одобрили окончательный вариант рукописи.
Авторы заявляют об отсутствии конфликта интересов.
Вклад соавторов: *Каждый автор внес равную долю участия, как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.*

Analysis of the Generation and Conversion Efficiency of Optical ILCF-signals on the Basis of a Classical RoF-segment Based on the Application of Two Lasers

Irina L. Vinogradova[1], Evgenia Yu. Golovina[2]

[1] FGBOU VO "Ufa University of Science and Technology"
[2] Institute of Oil Refining and Petrochemistry FSBEU VO "UGNTU" in Salavat city

Abstract: The paper proposes a method of generation and conversion of optical ILCF signals on the basis of a classical RoF-segment based on the application of two lasers. The circuit parameters of the classical RoF-segment based on the application of two independent laser sources are estimated. An approach to modeling the parameters of propagating radiation is proposed. It is found that in case of absence of signals chirping this scheme can be applied to the problems of radio frequency reception within the range up to several GHz. As the optical frequency distance of these lasers increases, the level of instability grows significantly negating the advantages of optical radio frequency generation due to increasing phase noise. Nevertheless, the range of parameters is defined (in terms of the difference in laser frequencies and the largest permissible value of ILCF) when the considered scheme provides a more or less stabilized mode of operation, and therefore it can find application in the tasks of the industrial internet.

Keywords: Radio-over-Fiber, fiber-optic transmission lines, frequency range, laser diodes, Mach–Zehnder modulator, industrial internet, Maple 12 software package, ILCF signals, chirping, phase noise.

For citation: Vinogradova I. L., Golovina E. Yu. Analysis of the generation and conversion efficiency of optical ILCF- signals on the basis of a classical RoF-segment based on the application of two lasers (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 3, pp. 28-44. https://doi.org/10.55648/1998-6920-2024-18-3-28-44.

Content is available under the license Creative Commons Attribution 4.0 License

© Vinogradova I. L., Golovina E. Yu., 2024

The article was submitted: 14.12.2023;
revised version: 31.01.2024;
accepted for publication 06.03.2024.

References

1. Vinogradova I. L., Golovina E. Yu. Ispolzovanie tehnologii Radio-over-Fiber v sistemah svyazi [Use of Radio-over-Fiber technology in communication systems]. *Collection of scientific articles on the materials of the XIII International Scientific and Practical Conference "Actual problems of science and technology"*, Ufa, 24 November, 2023, pp. 84-88.
2. Pham Van Dai. *Razrabotka modelei i metodov marshrytizacii v energoeffektivnih yacheistih setyah dal'nego radiusa deistviya* [Development of routing models and methods in energy-efficient long-range mesh networks]. Abstract of Ph. D. thesis. Saint Petersburg, 2021. 160 p.
3. Solomon Tesfay Abraha. *Impulse Radio Ultra Wideband over Fiber Techniques for Broadband In-Building Network Application*. Abstract Ph. D. thesis. Netherlands. 2012. 192p.
4. Haymen Shams. *Radio over Fiber Distribution Systems for Ultra Wideband and Millimetre wave Applicatio*ns. Abstract Ph. D. thesis. Dublin. 2011. 158p.
5. Atsushi Kanno, Pham Tien Dat, Toshiaki Kuri, Iwao Hosako, Tetsuya Kawanishi, Yuki Yoshida, and Ken'ichi Kitayama. Evaluation of frequency fluctuation in fiberwireless link with direct IQ down-converter. *European Conference on Optical Communication (ECOC)*, 2014, pp. 1-3.
6. U. Gliese, S. Norskov, and T. N. Nielsen. Chromatic dispersion in fiber-optic microwave and millimeter-wave links. *IEEE Transactions on microwave theory and techniques*, 1996, vol. 44, no. 10, pp. 1716-1724.
7. Recommendation ITU-R SM.1045-1. Frequency tolerance of transmitters. *Electronic Publication*, 1997, 5p.
8. Orazio Svelto. *Principles of Lasers*. Springer Science & Business Media, 2010, 620 p.
9. Govind P. Agrawal. *Fiber-optic communication systems with cd*. 4nd ed. Hoboken, New Jersey, John Wiley & Sons, Inc., 2010. 588 p.
10. Uekara N. and Ueda K. 193-mHz Beat Linewidth of Frequency Stabilized Laser-Diode-Pumped Nd:YAG Ring Lasers. *Opt. Lett*, 1 April, 1993, pp. 505-507. DOI: 10.1364/ol.18.000505.
11. Richard Scheps. *Introduction to Diode-Pumped Solid-State Lasers*. Lightwave Electronics Corp. Techn. Information, 2003. 120 p.
12. Ikegami T., Sudo S. and Sakai Y. *Frequency Stabilization of Semiconductor Laser Diodes*. Boston, Artec House, 1995. 356 p.
13. Guy M., Villeneuve B., Latrasse C. and Tètu M., Lightwave J. Simultaneous absolute frequency control of laser transmitters in both1.3 and 1.55 mu-m bands for multiwavelength communication-systems. *Journal of lightwave technology*, 1996, pp. 1136-1143.
14. Lee Hyunjae, Lyu, G. Y., Park S. Y. and Lee J. H. Multichannel wavelength locking using transmission peaks of an AWG for multichannel optical transmission systems. *IEEE Photonics Technology Letters*, 1998, vol. 10, no. 2, pp. 276-278.
15. Park Y., Lee S. T. and Chae C. J. A novel wavelength stabilization scheme using a fiber grating for WDM transmission. *IEEE Photon. Technology Letters*, 1998, vol. 10, no. 10, pp. 1446–1448.
16. Ono T. and Yano Y. Key technologies for terabit/second WDM systems with high spectral efficiency of over 1 bit/s/Hz. *IEEE J. Quantum Electron*, 1998, vol. 34, pp. 2080-2088.
17. Nasu H., Takagi T., Shinagawa T., Oike M., Nomura T. and Kasukawa A., Lightwave J. A Highly Stable and Reliable Wavelength Monitor Integrated Laser Module Design. *Journal Lightwave Technology*, 2004, vol. 22, pp. 1344-1351.
18. Nasu H., Mukaihara T., Takagi T., Oike M., Nomura T. and Kasukawa A. Wavelength monitor integrated laser modules for 25-GHz-spacing tunable applications. *IEEE Journal of Selected Topics in Quantum Electronics*, 2005, vol. 11, no. 1, pp. 157-164.
19. Tissot Y., Limberger H. G. and Salathe R.-P. Ultrawide bandwidth wavelength monitor based on a pair of tilted fiber Bragg gratings. *IEEE Photon. Technol. Lett*, 2007, vol. 19, pp. 1702-1704.
20. Liang W., Ilchenko V. S., Savchenkov A. A., Matsko A. B., Seidel D. and Maleki L. Whispering gallery mode resonator based ultra-narrow linewidth external cavity semiconductor laser. *Optics letters*, 2010, vol. 35, pp. 2 822-2824.
21. *Thorlabs. Models: EM650 High Power DFB Laser Module; SFL1550S*, available at: `https://www.thorlabs.com/` (accessed: 11.09.2023)
22. *Model RIO PLANEXSeries 1550nm Low Phase Noise Narrow Linewidth External Cavity Laser, Redfern Integrated Optics (RIO)*, available at: `http://rio-lasers.com/` (accessed: 01.10.2023).
23. Consolino Luigi, Nafa Malik, Cappelli Francesco Garrasi Katia, Mezzapesa Francesco P., Li Lianhe, Davies A. Giles, Linfield Edmund H., Vitiello Miriam S., De Natale1 Paolo and Bartalini Saverio. Fully

phase-stabilized quantum cascade laser frequency comb. *Electronic and Electrical Engineering*, 2018, vol. 2, no. 5, pp. 231-237.
24. Lang R., Kobayashi K. External optical feedback effects on semiconductor injection laser properties. *EEE J. Quantum Electron*, 1980, pp. 347–355.
25. Javaloyes J., Mandel P., Pieroux D. Dynamical properties of lasers coupled face to face. *Physical review. Statistical, nonlinear, and soft matter physics*, 2003, pp. 11.
26. Burdin Vladimir A., Dashkov Mikhail V., Volkov Kirill A. Application of dispersion managed soliton regime in radio-over-fiber systems. *Proceedings of SPIE*, 2012, vol. 8787, pp. 1-6.
27. Abdrakhmanova G.I., Andrianova A.V., Vinogradova I.L., Yantilina L.Z., Grahova E.P., Ishmiyarov A.A., Khasanshin V.R., Meshkov I.K., Sultanov A.Kh. Interference fiber optic device for RoF antenna radiators control. *International Siberian Conference on Control and Communications*, 2016, pp.1-6.
28. Agrawal G. P. *Nonlinear fiber optics*. Academic Press, Boston, 2009. 466p.
29. Chorchos Łukasz, Rommel Simon, Turkiewicz Jarosław P., Monroy Idelfonso Tafur and Olmos Juan Jos´e Vegas. *Reconfigurable Radio Access Unit for DWDM to W-Band Wireless Conversion. IEEE Photonics technology let*, available at: http://www.ieee.org/publications_standards/publications/rights/index.html/ (accessed: 08.10.2023).
30. Meijerink Arjan, Roeloffzen Chris, Meijerink Roland, Zhuang Leimeng, Marpaung David, Bentum Mark J., Burla Maurizio, Verpoorte Jaco, Jorna Pieter, Hulzinga Adriaan and van Etten Wim. Novel Ring Resonator-Based Integrated Photonic Beamformer for Broadband Phased Array Receive Antennas - Part I: Design and Performance Analysis. *J. lightwave technology*, January 2010, vol. 28, no. 1. pp. 3–18.
31. Leimeng Zhuang, Chris G. H. Roeloffzen, Arjan Meijerink, Maurizio Burla, David A. I. Marpaung, Arne Leinse, Marcel Hoekman, René G. Heideman, and Wim van Etten. Novel Ring Resonator-Based Integrated Photonic Beamformer for Broadband Phased Array Receive Antennas - Part II: Experimental Prototype. *Journal of lightwave technology*, January 2010, vol. 28, no. 1, pp. 19–31.
32. Alphones Arokiaswami and Quang Thai Pham. Hybrid Approach for Optical Beamforming for Phased Array. *Proceedings of Asia-Pacific Microwave Conference*, 2010, pp. 311–317.
33. Drummond Miguel V., Monteiro Paulo P. and Nogueira Rogério N. Photonic True-Time Delay Beamforming Based on Polarization-Domain Interferometers. *J. Lightwave technology*, 2010, vol. 28, no. 17, pp. 2492 – 2498.
34. Burla M., Khan M. R. H., Marpaung D. A. I., Roeloffzen C. G. H., Maat P., Dijkstra K., Leinse A., Hoekman M., Heideman R. Squint-Free Beamsteering Demonstration using a Photonic Integrated Beamformer based on Optical Ring Resonators. *IEEE International Topical Meeting on Microwave Photonics*, 2010, pp. 401-404.
35. Kivshar Y. S., Agrawal G. P. *Opticheskie solitoni. Ot volokonnih svetovodov do fotonnih kristallov* [Optical solitons. From fiber-optic light guides to photon crystals]. Transl. from English, ed. by N. N. Rozanov. Moscow, Fizmatlit, 2005, 648 p.

Vinogradova Irina L.

Doctor of Technical Sciences, Professor, Department of Telecommunication Systems, Ufa University of Science and Technology (UUNiT, 32, Zaki Validi St., Ufa, 450076, Republic of Bashkortostan, Ufa, 450076), e-mail: vil-4@mail.ru, ORCID ID: 0000-0003-4618-6843.

Golovina Evgenia Yu.

Senior Lecturer, Department of Information Technologies, Institute of Oil Refining and Petrochemistry, FSBEU VO "UGNTU" in Salavat (Institute of Oil Refining and Petrochemistry FSGBOU VO "UGNTU" in Salavat, 453250, Republic of Bashkortostan, Salavat, Gubkina St., 22B), e-mail: egolovina82@mail.ru, ORCID ID: 0009-0006-3624-5490.

DOI: 10.55648/1998-6920-2024-18-3-45-56 УДК 519.718

Одна задача структурной оптимизации сетей передачи данных *

А. М. Кальней[1], А. С. Родионов[1,2]

[1] Институт вычислительной математики и математической геофизики СО РАН (Новосибирск)
[2] Сибирский гос. унив. телекоммуникаций и информатики (СибГУТИ)

Аннотация: В статье рассматриваются вопросы анализа и структурной оптимизации многоуровневых сетей передачи данных с использованием модели случайной гиперсети. Утверждается, что основным уровнем, подверженным разрушениям под воздействием внешней среды и влияющим на работоспособность сети в целом, является физический уровень, но в структурном плане физические разрушения прежде всего сказываются на уровне канальном. В свою очередь, отказы на канальном уровне влияют на организацию работы сетевого и вышестоящих уровней, в частности, требуют перераспределения потоков, в том числе перестроения виртуальных каналов.

Приводится краткий обзор моделей многоуровневых сетей, используемых для анализа и структурной оптимизации сетей передачи данных различного назначения (в частности, сенсорных сетей мониторинга), даётся описание случайной гиперсети и обсуждаются некоторые задачи повышения надёжности функционирования сетей на сетевом уровне при возможных отказах каналов и узлов сети, т.е. нарушениях её работы на канальном уровне.

В качестве показателей надёжности сетевого уровня сети рассматриваются возможность (вероятность) передачи данных между парой узлов в принципе либо по установленным виртуальным каналам и математическое ожидание числа узлов, способных передать данные в центральный узел.

В качестве целей структурной оптимизации рассматриваются наилучшая укладка виртуальных каналов в известную структуру канальной сети и размещение базовых станций в узлах сети, связь которых с центральным узлом считается надёжной.

Ключевые слова: надежность сети, гиперсети, структурная оптимизация.

Для цитирования: Кальней А. М., Родионов А. С. Одна задача структурной оптимизации сетей передачи данных // Вестник СибГУТИ. 2024. Т. 18, № 3. С. 45–56. https://doi.org/10.55648/1998-6920-2024-18-3-45-56.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Кальней А. М., Родионов А. С., 2024

Статья поступила в редакцию 22.01.2024;
переработанный вариант – 12.03.2024;
принята к публикации 14.03.2024.

1. Введение

Сети связи вообще и сети передачи данных в частности относятся к классу многоуровневых сетей. Одной из наиболее известных моделей таких сетей является модель взаимодействия

*Данные работы выполнены по государственному контракту с ИВМиМГ СО РАН № 0251-2022-0005.

открытых систем OSI (Open Systems Interconnection). Имеем решающее влияние на функционирование таких сетей уровня канального, причины неполадок которого чаще всего обусловлены воздействием внешней среды на уровень физический. Также не рассматриваем и уровни выше сетевого, надёжность которых определяется аппаратно-программной реализацией соответствующих протоколов и работа которых возможна только в условиях достаточного уровня надёжности канального и сетевого уровней.

Как в связи с рассмотрением цифровых сетей, так и сетей другой природы (транспортных, социальных, информационных и др.) возникает потребность построения адекватных моделей, их описывающих. В работе [1] дан критический обзор некоторых моделей. В частности, для телекоммуникационных сетей исследованы следующие модели:

- Сэндвич-графы [2].
- Модель многоуровневой сети, основанной на ITU-T G.805 [3].
- Модель для проектирования многоуровневой структуры WLAN [4].
- Многоуровневые сложные сети (LCN) [5].
- Модель для двухуровневой SDH/WDM-сети [6].
- Модель для двухуровневой оптической сети [7].
- Многослойная сеть [8].
- Гиперсеть [9].

Основной задачей любого информационного управления в сети является анализ поведения сети и моделирование вариантов ее развития на основе реальной информации. Чем сложнее топология и конфигурация сети, тем больше информации требуется для проведения ее адекватного анализа. При анализе или проектировании больших сетей мониторинга часто возникает проблема расположения контрольных устройств (b-узлов) для сбора информации. После некоторой предварительной обработки или напрямую b-узлы передают информацию центральному узлу (c-узлу) по надежным каналам. Одним из основных показателей качества таких сетей является размер территории, находящейся под надежным контролем, которую можно оценить с помощью математического ожидания количества узлов, связанных с одним специальным узлом MENC (от англ. Mathematical expectation of a number of nodes that are connected with one special node).

Остальная часть статьи организована следующим образом: во второй главе рассматриваются вопросы применимости исследованных моделей для анализа и оптимизации телекоммуникационных сетей, в третьей главе вводится модель гиперсети, в четвёртой – описывается задача структурной оптимизации многоуровневой сети, в пятой – приводится пример.

2. Применимость моделей для телекоммуникационных сетей

Широко используемая на практике модель графов не подходит для моделирования даже простейшей двухуровневой сети, что показывает следующий пример. Пусть имеется кабельная сеть, представленная графом G_1, и пусть некоторая сеть, осуществляющая транспортировку данных, лежит в нём. Эта сеть может быть представлена графом G_2, который необязательно совпадает с G_1. Назовём G_1 физической сетью, в то время как G_2 назовём логической сетью. Вложение G_2 в G_1 может быть сделано различными способами. Имеется некоторое отображение связей G_2 в пути, состоящие из связей G_1. Очевидно, выход из строя или изменение пропускной способности ребра графа G_1 может привести к выходу из строя или изменению пропускной способности нескольких связей в G_2 или не изменить ни одну из них. Поэтому для анализа многоуровневой сети необходима более сложная модель, чем граф.

Относительно недавно были проведены серьёзные исследования моделей многоуровневых сложных сетей, обобщения терминологии и инструментов в этой области [8, 10, 11]. В работе

[8] была введена модель многослойной сети, которая может использоваться для представления большинства типов сложных систем (например, в социологии, эпидемиологии, биомедицине и т.д.), которые состоят из нескольких сетей или включают несопоставимые и/или множественные взаимодействия между объектами. Более подробное описание этой модели и логики её построения приведено в работе [1], также в этой работе приведено сравнение с моделью гиперсети, показавшее большую универсальность последней модели. Ниже даётся краткое описание упомянутых моделей многоуровневых сетей.

2.1. Модель многоуровневой сети, основанной на ITU-T G.805

В работе [3] рассматривается задача маршрутизации в сети передачи данных с коммутацией каналов с наличием нескольких логических слоёв (разных технологий) и различных интерфейсов (с возможностью адаптации сигнала той или иной технологии). В таких сетях естественным образом возникает задача проверки валидности устанавливаемого пути между узлами. На рис. 1 из работы представлен пример такой сети, где каждый узел является национальной исследовательской, образовательной сетью, участвующей в глобальном интегрированном фонде Lambda. Связи между ними могут быть: соединение Gigabit Ethernet (GE), соединения OC-192, основанные на технологии SONET (Synchronous Optical Networking) и передающие 192 канала STS (Synchronous Transport Signal).

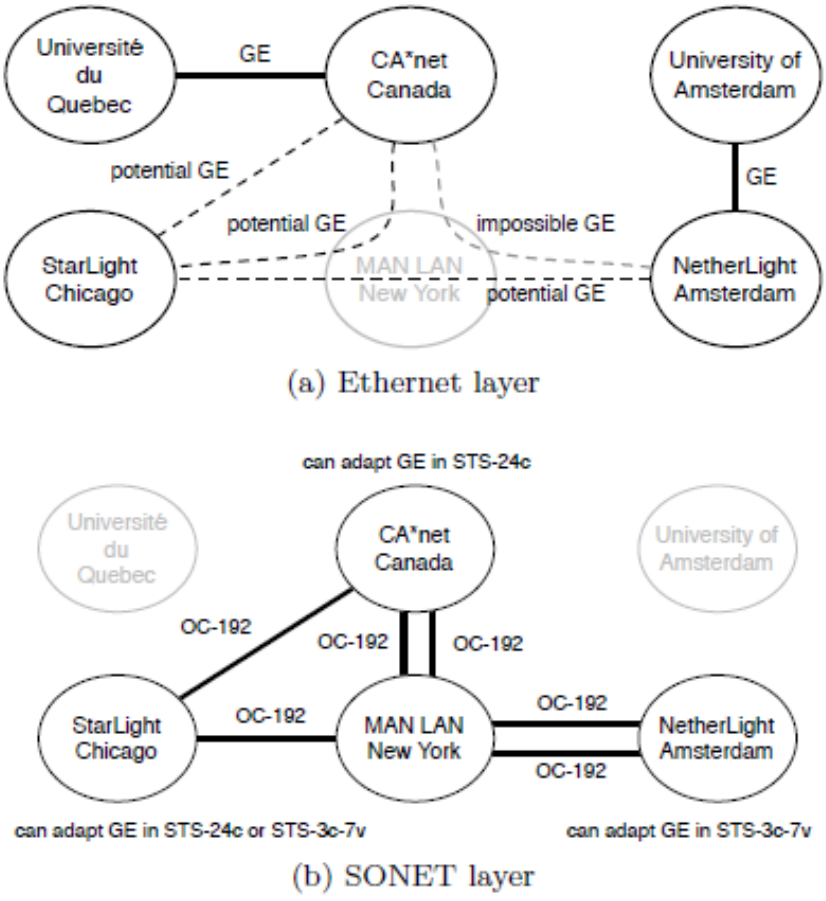


Рис. 1. Пример двухуровневой сети с коммутацией каналов

В работе показано, что из-за различий в возможности адаптации соединения GE невозможно установление соединения по кратчайшему пути между университетом Квебека (Université du

Quebec) и университетом Амстердама (University of Amsterdam) через MAN (Metropolitan Area Network) LAN (Local Area Network) New York. Как заключают авторы статьи, многоуровневый алгоритм поиска пути должен иметь информацию не только об уровнях и адаптациях своих непосредственных соседей, но и об уровнях и адаптациях доменов, расположенных дальше по пути. Для моделирования таких сетей было предложено использовать рекомендации ITU-T (International Telecommunication Union — Telecommunication sector) G.805 [12]. Основываясь на данных рекомендациях, было проведено отображение элементов сети в функциональные элементы, отображённые в табл. 1. Авторами была введена простая алгебра для нахождения валидного пути между двумя узлами.

Т а б л и ц а 1. Отображение элементов сети с коммутацией каналов в функциональные элементы

Элемент сети	Функциональные элементы
Домен	Подсеть(и)
Устройство	Матрица (Подсеть)
Интерфейс	Точка(и) подключения и функция(и) адаптации
Связь	Соединение по связи

Однако применительно к задаче анализа надёжности многоуровневых сетей мониторинга различной природы данная модель не может быть применена, поскольку она ограничивается только сетями с коммутацией каналов.

2.2. Модели для прикладных задач

Зачастую при решении конкретной прикладной задачи исследователи создают частные модели для их решения. В качестве примера таких задач можно привести проектирование многоуровневой структуры WLAN (Wireless Local Area Network) [4], двухуровневой сети SDH/WDM (Synchronous Digital Hierarchy/Wavelength-Division Multiplexing) [6], разработку схемы защиты (восстановления) двухуровневой оптической сети [7].

В задаче проектирования многоуровневой структуры WLAN решаются проблемы качества обслуживания, безопасности и мобильности (возможности перемещения пользователей между областями) сети. Для решения данной задачи используется межуровневое проектирование в сети, основанной на протоколе TCP/IP (Transmission Control Protocol/Internet Protocol), состоящей из пяти уровней: прикладной уровень, транспортный уровень, сетевой уровень, уровень канала передачи данных и физический уровень. Для достижения каждой поставленной задачи вводятся вертикальные связи между уровнями сети.

При проектировании двухуровневой SDH/WDM-сети ставятся цели найти минимальную стоимость установки аппаратного оборудования каналов и узлов на обоих сетевых уровнях, чтобы требования к трафику могли быть реализованы посредством регулируемой (процесс группировки множества мелких телекоммуникационных потоков в более крупные) и устойчивой маршрутизации. Для решения данной задачи вводится формулировка смешанно-целочисленного программирования для предопределенного набора допустимых логических каналов, которая учитывает множество практических ограничений, включая аппаратное обеспечение узла, несколько скоростей передачи данных и устойчивость к отказам одного физического узла или канала.

В двухуровневой оптической сети преобладают следующие два типа отказов: сбои на оптическом транспортном уровне, такие как обрыв волокна или кабеля или отказ передатчика, и сбои на пакетном уровне: неправильная конфигурация, сбой процессора или линейной карты, сбои питания или сбой источника питания. Используя для пакетного уровня сеть на основе протокола IP/MPLS (Internet Protocol/Multiprotocol Label Switching), а также известные механизмы

для восстановления элементов сети как на оптическом, так и на пакетном уровне, достигаются задачи повышения эффективности сети в целом. Для изучения предложенного подхода была разработана модель смешанно-целочисленного программирования, решение которого минимизирует общую стоимость двухуровневой оптической сети для защиты трафика от отказов.

2.3. Многоуровневые сложные сети (LCN)

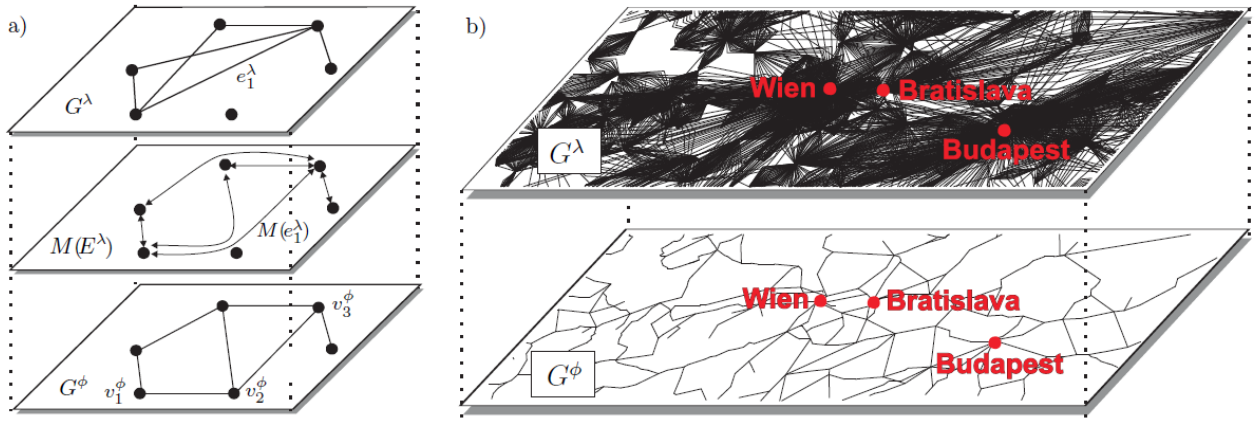


Рис. 2. (a) Иллюстрация двухуровневой модели. Логический граф G^λ отображается в физический граф G^ϕ по отображению $M(E^\lambda)$. В этом примере логическое ребро e_1^λ отображается в граф G^ϕ как путь $M(e_1^\lambda) = (v^{\phi_1}), v^{\phi_2}, v^{\phi_3}$. Предполагается, что G^λ – невзвешенный, нагрузки трёх указанных узлов равны: $l(v_1^\phi) = 3$, $l(v_2^\phi) = 2$, $l(v_3^\phi) = 4$. (b) Часть логических и физических графов для набора данных Европы. Здесь интенсивности трафика (веса) обозначаются мультиребрами логического графа.

Для оценки распределения трафика вводится двухуровневая модель, состоящая из физического и логического уровней [5]. Во многом эта модель схожа с вводимой в следующей главе моделью гиперсети. Как и для гиперсети, вводится физический граф (первичная сеть в терминологии гиперсети) $G^\phi = (V^\phi, E^\phi)$ и логический граф (вторичная сеть в терминологии гиперсети) $G^\lambda = (V^\lambda, E^\lambda)$. Как и для простой гиперсети, предполагается, что вершины физического и логического графов совпадают: $N = V^\phi = V^\lambda$. Каждое логическое ребро $e^\lambda = (u^\lambda, v^\lambda)$ отображается в маршрут на физическом графе $M(e^\lambda) \subset G^\phi$. Узлам и ребрам могут быть присвоены веса, обозначенные $w(\cdot)$, причем $w = 1$ для невзвешенных графов. Нагрузкой l узла v^ϕ является сумма всех весов логических рёбер, чьи пути проходят через v^ϕ:

$$l(v^\phi) = \sum_{e^\lambda : v^\phi \in M(e^\lambda)} w(e^\lambda). \tag{1}$$

На картинке 2, взятой из статьи (рис. 2b) , представлен пример двухуровневой транспортной сети (примерами таких сетей могут быть сети энергосистемы, железнодорожные сети, дорожные сети, городские системы общественного транспорта). Также на картинке можно видеть способ укладки логических рёбер в физический граф.

Однако, несмотря на схожесть этой модели с моделью гиперсети, она учитывает связи только соседних уровней (физического и логического). В более сложных сетях может быть больше двух уровней, на одном уровне может быть несколько сетей, а вершины могут не совпадать. Для моделирования таких сетей необходимо использовать более сложную модель S-гиперсети [13].

3. Модель гиперсети

Перейдём к описанию гиперсетевых моделей. Двухуровневая, или простая, гиперсеть $H = (X, V, R; P, F, W)$ состоит из следующих объектов (см. рис. 3, заимствованный из [14]):

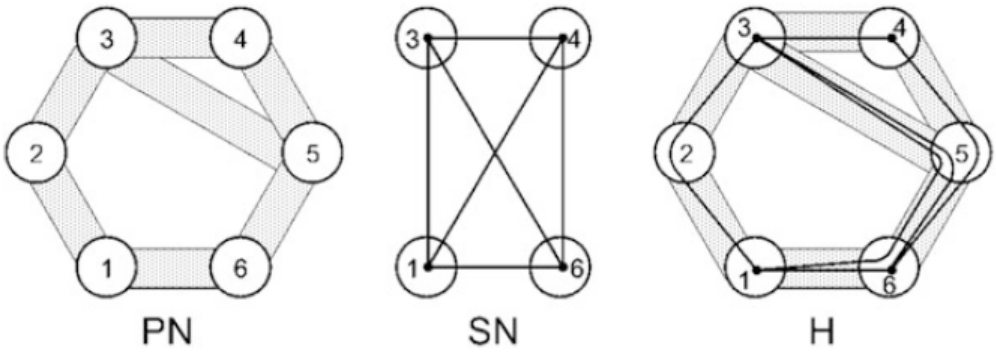


Рис. 3. Иллюстрация к определению гиперсети: PN – это первичная сеть; тёмные вершины $\{1, 3, 4, 6\}$ из набора вершин относятся к SN; SN – это вторичная сеть (в нашем случае является полным графом); H – это гиперсеть (SN отображается в PN)

$X = (x_1, \dots, x_n)$ – набор вершин;
$V = (v_1, \dots, v_m)$ – набор ветвей (рёбер графа первичной сети);
$R = (r_1, \dots, r_g)$ – набор рёбер (рёбер графа вторичной сети);
$P : V \to X \times X$ – отображение, которое определяет граф $PN = (X, V)$, который называется первичной сетью;
$W : R \to X \times X$ – отображение, которое определяет граф $SN = (X, R)$, который называется вторичной сетью;
$F : R \to 2^V$ – отображение рёбер в маршруты PN.

Вводя множественное вложение вторичных сетей в первичную сеть, вложение вторичных сетей друг в друга, отображение нескольких вершин вторичной сети в узел первичной, приходим к формальному определению S-гиперсети [13].

4. Структурная оптимизация гиперсети

Возможны различные разрушения гиперсети, они изображены на рис. 4, заимствованном из [14]. В текущей работе используются варианты I, III и IV для удаления узлов, ветвей и рёбер соответственно.

Первым рассматриваемым показателем надежности будет вероятность соединения пары узлов во вторичной сети R_{st}. Выбранные узлы обозначаются как s, t.

Одним из возможных аддитивных показателей, который можно получить из этого показателя, является $MENC$, который можно рассчитать по формуле:

$$\sum_{1 \le i \le n} R_{CNx_i}(H).$$

Центральный узел помечен как CN. Этот показатель используется как характеристика качества сети мониторинга с точки зрения охвата контролируемой территории.

В работе [15] был представлен метод факторизации для вычисления парной связности R_{st}, были описаны различные приёмы по ускорению его вычисления. При вычислении $MENC$ мы получаем кумулятивные границы [16], которые могут помочь остановить вычисления, когда требуемое значение $MENC$ больше верхней границы или меньше нижней границы.

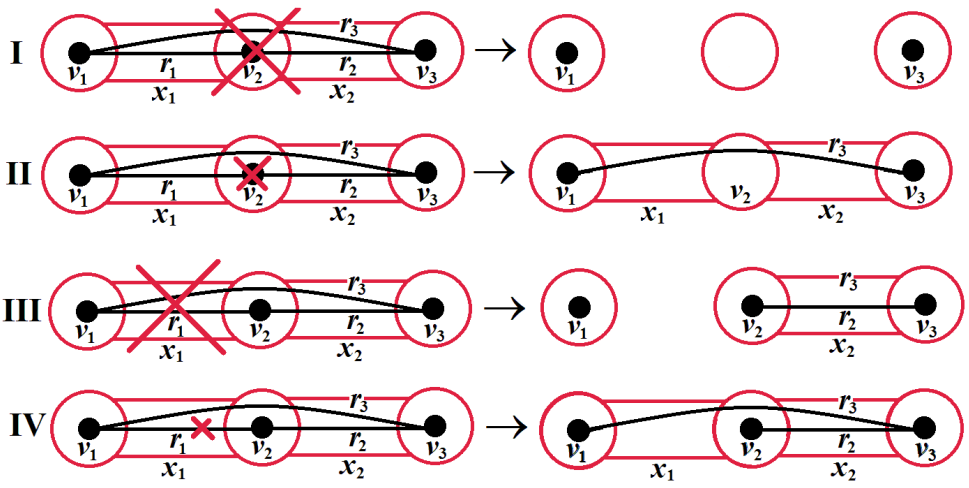


Рис. 4. Возможные случаи разрушения гиперсети

Рассмотрена классическая модель «затраты – надежность»:

$$\sum_{j\in 1,\ldots,m} c_j y_j \to \min_{y_j}, \tag{2}$$

$$MENC \geq R, \tag{3}$$

где c_j – стоимость размещения b-узла в ветви v_j; y_j – флаг выбранной ветви v_j; R – требуемое значение надежности. Если все затраты равны, то получаем следующую задачу:

$$\sum_{j\in 1,\ldots,m} y_j \to \min_{y_j},$$

$$MENC \geq R.$$

Каждое размещение b-узла на ветви изменяет элементы гиперсети, как показано на рис. 5. Сначала мы получаем b-узел, который разделяет ветвь и создает новые рёбра во вторичной сети. Затем соединяем его надежной ветвью с c-узлом.

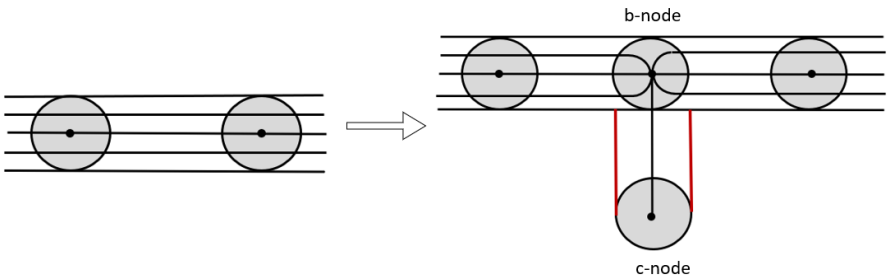


Рис. 5. b-узел на ветви

Задача точного расчета вероятности связности случайной сети с ненадежными элементами относится к классу NP-трудных задач [17]. Поэтому были рассмотрены различные эвристические алгоритмы. Алгоритм имитации отжига (ИО) кажется вполне подходящим, поскольку он не требует слишком большого вычисления $MENC$ в условии (3). На каждом шаге алгоритма ИО мы находим решение, удовлетворяющее условию (3), путем случайной замены одной точки. Проверенные решения сохраняются в памяти.

5. Пример

Рассмотрим задачу проектирования гиперсети с пропускной способностью каналов и максимальной длиной маршрутов. Для решения этой задачи ищем кратчайший маршрут для каждой пары узлов из X на ветвях первичной сети. Используется алгоритм поиска в ширину. Решая эту задачу, вкладываем вторичную сеть в первичную и, соответственно, получаем гиперсеть.

В качестве исходных данных рассмотрим решетку размером 3 на 3 $(Grid_{3\times3})$ (на данной размерности можно легко проверить вычисления вручную) с пропускной способностью для всех каналов, равной двум, и максимальной длиной маршрутов, равной шести, представленную на рис. 6. Красным цветом заданы стоимости расстановки контрольных устройств на ветвях. В качестве порога надёжности примем значение *MENC* $R = 7.8$.

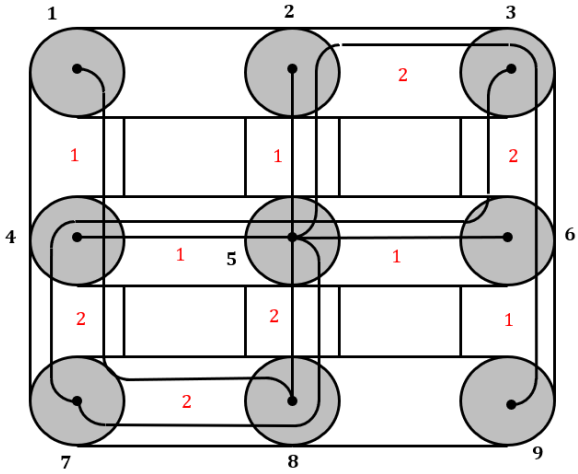


Рис. 6. Демонстрационная гиперсеть

Рассмотрим решение $(4,7);(5,6)$, представленное на рис. 7. Данное решение имеет значение целевой функции, равное 3, а значение *MENC* – 7.81924.

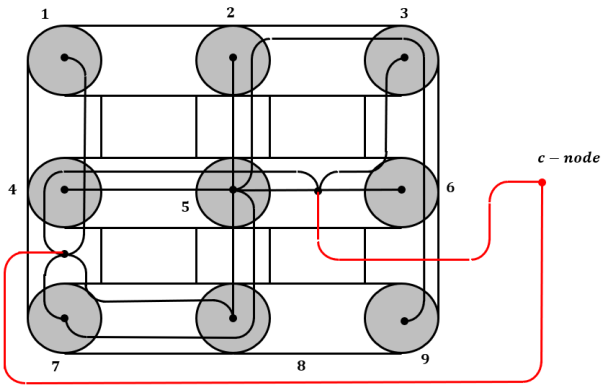


Рис. 7. Решение, полученное с помощью алгоритма ИО

В качестве негативного примера можно привести решение $(4,5);(5,6)$, представленное на рис. 8. Хотя данное решение имеет значение целевой функции, равное 2, оно не удовлетворяет требованию надёжности (3), поскольку имеет значение *MENC*, равное 7.67268.

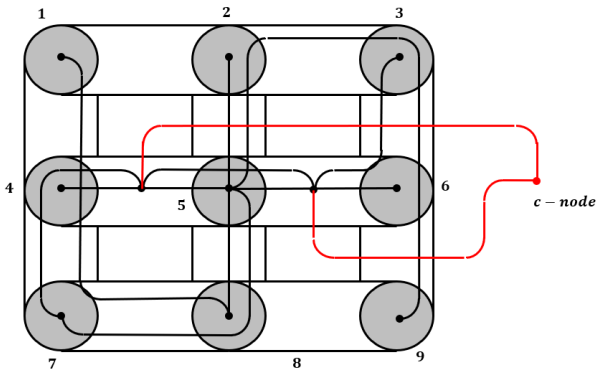


Рис. 8. Негативный пример

6. Заключение

В статье приведён краткий обзор моделей, используемых для проектирования многоуровневых телекоммуникационных сетей. Для задачи повышения надёжности функционирования многоуровневых сетей передачи данных была выбрана модель случайной гиперсети. В качестве показателей надёжности сетевого уровня сети рассматриваются возможность (вероятность) передачи данных между парой узлов в принципе либо по установленным виртуальным каналам и математическое ожидание числа узлов, способных передать данные в центральный узел. Была описана формулировка задачи структурной оптимизации двухуровневой сети мониторинга с ненадёжными каналами (ветвями или рёбрами первичной сети в терминологии гиперсетей). Приведён пример работы алгоритма имитации отжига на небольшой гиперсети.

Литература

1. *Кальней А. М.* Модели многоуровневых сетей (краткий обзор) // Проблемы информатики. 2021. № 3. С. 5 – 20.
2. *Kim J. and Vu V.* Sandwiching random graphs: Universality between random graph models // Advances in Mathematics. 2004. V. 188, № 2. P. 444 – 469.
3. *Dijkstra F., Andree B., Koymans K., van der Hama J., Grosso P., de Laat C.* A multi-layer network model based on ITU-T G.805 // Computer Networks. 2008. V. 52, № 7. P. 1927 – 1937.
4. *Rotaru S.*. Cross-layer design in wireless local area networks (WLANs) issues and possible solutions. 2018. № 12.
5. *Kurant M., Thiran P.* Layered complex networks // Physical review letters. 2006. V. 96, № 5. P. 138701.
6. *Orlowski S., Raack C., Koster A., Baier G., Engel T., Belotti P.* Branch-and-cut techniques for solving realistic two-layer network design problems // Graphs and Algorithms in Communication Networks. 2010. № 1. P. 95 -– 118.
7. *Chigan C., Atkinson G., Nagarajan R.* On the modeling issue of joint cross-layer network protection/restoration // Proceedings of Advanced Simulation Technologies Conference (ASTC ’04). 2004. P. 57 -– 62.
8. *Kivela M., Arenas A., Barthelemy M., Gleeson J. P., Moreno Y. and Porter M. A.* Multilayer networks // Journal of Complex Networks. 2013. V. 2, № 3. P. 203 – 271. URL: `https://doi.org/10.1093/comnet/cnu016`.

9. *Popkov V. K., Sokolova O. D.* Application of Hyperneet Theory for the Networks Optimazation Problems // 17th IMACS World Congress, Scientific Computation, Applied Mathematics and Simulation. Paris. 11 – 15 July, 2005. Paper T4-I-42-0112.
10. *Boccaletti S., Bianconi G., Criado R., del Genio C. I., Gomez-Gardenes J., Romance M., Sendina-Nadal I., Wang Z., & Zanin M.* The structure and dynamics of multilayer networks // Physics Reports. 2014. V. 544, № 11. P. 1 – 122.
11. *Domenico M. D., Granell C., Porter M. A., & Arenas A.* The physics of spreading processes in multilayer networks // Nature Physics. 2016. V. 12. P. 901 – 906.
12. ITU-T Recommendation G.805: Generic functional architecture of transport networks // Tech. rep., International Telecommunication Union. March 2000. URL: http://www.itu.int/rec/T-REC-G.
13. *Попков В. К.* Применение теории S-гиперсетей для моделирования систем сетевой структуры // Проблемы информатики. 2010. № 4. С. 17 – 40.
14. *Rodionov A., Rodionova O.* Random Hypernets in Reliability Analysis of Multilayer Networks // Electrical Engineering. 2015. V. 343, № 1. P. 307 – 315. URL: https://doi.org/10.1007/978-3-319-15765-8-17.
15. *Кальней А. М.* Оптимизация размещения контрольных устройств на каналах в сетях мониторинга // Проблемы информатики. 2022. № 4. С. 28 – 38. DOI: 10.24412/2073-0667-2022-4-28-38.
16. *Rodionov A. S., Migov D. A.* Obtaining and Using Cumulative Bounds of Network Reliability // System reliability. Rijeka: IntechOpen, 2017. P. 93 – 112. DOI: 10.5772/intechopen.72182.
17. *Valiant L. G.* The complexity of computing the permanent // Theoretical Computer Science. 1979. V. 8, № 2. P. 189 – 201.

Кальней Артём Максимович

инженер, лаборатория системного моделирования и оптимизации, Институт вычислительной математики и математической геофизики СО РАН (ИВМиМГ СО РАН, 630090, Новосибирск, ул. Проспект Академика Лаврентьева, д. 6), e-mail: artemkalnei@yandex.ru, ORCID ID: 0009-0003-4107-7995.

Родионов Алексей Сергеевич

д.т.н., заведующий лабораторией системного моделирования и оптимизации, Институт вычислительной математики и математической геофизики СО РАН (ИВМиМГ СО РАН, 630090, Новосибирск, ул. Проспект Академика Лаврентьева, д. 6); профессор кафедры вычислительных систем, Сибирский государственный университет телекоммуникаций и информатики (СибГУТИ), тел. +7 383 3326949, e-mail: alrod@sscc.ru, ORCID ID: 0000-0001-8411-4707.

Авторы прочитали и одобрили окончательный вариант рукописи.

Авторы заявляют об отсутствии конфликта интересов.

Вклад соавторов: *Каждый автор внёс равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.*

One Problem of Structural Optimization of Data Transmission Networks

Artyom M. Kalney [1], **Aleksej S. Rodionov** [1,2]

[1] Institute of Computational Mathematics and Mathematical Geophysics SB RAS (Novosibirsk)
[2] Siberian State University of Telecommunications and Information Science (SibSUTIS)

Abstract: The article discusses issues of analysis and structural optimization of multi-level data transmission networks using the random hypernetwork model. It is argued that the main level subjected to destruction under the influence of the external environment and affecting the performance of the network as a whole is the physical level, but in structural terms physical destruction primarily affects the channel level. In turn, failures at the link level affect the organization of work of the network and higher levels; in particular, they require the redistribution of flows including the rebuilding of virtual channels.
A brief overview of multi-level network models used for the analysis and structural optimization of data transmission networks for various purposes (in particular sensor monitoring networks) is presented, a description of a random hypernetwork is given and some problems of increasing the reliability of the functioning of networks at the network level in the event of possible failures of channels and network nodes are discussed, that is, disruption of its operation at the link level.
As indicators of the reliability of the network layer of the network, the possibility (probability) of data transmission between a pair of nodes in principle or via established virtual channels and the mathematical expectation of the number of nodes capable of transmitting data to the central node are considered.
The goals of structural optimization are the best placement of virtual channels in the known structure of the channel network and the placement of base stations in network nodes which connection with the central node is considered to be reliable.

Keywords: network reliability, hypernets, structural optimization.

For citation: Kalney A. M., Rodionov A. S. One problem of structural optimization of data transmission networks (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 3, pp. 45-56.
https://doi.org/10.55648/1998-6920-2024-18-3-45-56.

Content is available under the license Creative Commons Attribution 4.0 License

© Kalney A. M., Rodionov A. S., 2024

The article was submitted: 22.01.2024;
revised version: 12.03.2024;
accepted for publication 14.03.2024.

References

1. Kalney A. M. Modeli mnogourovnevyh setej (kratkij obzor) [Multilayer network models (brief overview)]. *Problems of computer science*, 2021, no. 3, pp. 5-20.
2. Kim J., Vu V. H. Sandwiching random graphs: Universality between random graph models. *Advances in Mathematics*, 2004, vol. 188, pp. 444-469.
3. Dijkstra F., Andree B., Koymans K., Van der Ham J., Grosso P., and Laat C. A multi-layer network model based on ITU-T G.805. *Computer Networks*, 2008, vol. 52, pp. 1927-1937.
4. Rotaru S. *Cross-layer design in wireless local area networks (WLANs): issues and possible solutions*. 2018.
5. Kurant M., Thiran P. Layered Complex Networks. *Physical review letters*, 2006, vol. 96, pp. 138701.
6. Orlowski S., Raack C., Koster A., Baier G., Engel T., and Belotti P. *Branch-and-Cut Techniques for Solving Realistic Two-Layer Network Design Problems*, 2010, pp. 95-118.
7. Chigan C., Atkinson G., Nagarajan R. On the modeling issue of joint cross-layer network protection/restora-

tion. *Advanced Simulation Technologies Conference (ASTC '04)*, 2004, pp. 57-62.
8. Kivelä M., Arenas A., Barthelemy M., Gleeson J. P., Moreno Y., and Porter M. A. Multilayer networks. *Journal of Complex Networks*, 2013, vol. 2, pp. 203-271.
9. Popkov V. K., Sokolova O. D. Application of Hyperneet Theory for the Networks Optimazation Problems. *17th IMACS World Congress, Scientific Computation, Applied Mathematics and Simulation*, 2005.
10. Boccaletti S., Bianconi G., Herrero R., Genio C., Gómez-Gardeñes J., Romance M., Sendiña-Nadal I., Z.Wang, and Zanin M. The Structure and Dynamics of Multilayer Networks. *Physics Reports*, 2014, vol. 544, pp. 1-122.
11. Domenico M. D., Granell C., Porter M. A., and Arenas A. The physics of spreading processes in multilayer networks. *Nature Physics*, 2016, vol. 12, pp. 901-906.
12. ITU-T Recommendation G.805: Generic functional architecture of transport networks : Rep. *Tech. rep., International Telecommunication Union*, 2000.
13. Popkov V. K. Primenenie teorii S-gipersetej dlya modelirovaniya sistem setevoj struktury [Application of the theory of S-hypernets for modeling network structure systems]. *Problems of computer science*, no. 4, 2010, pp. 17-40.
14. Rodionov A., Rodionova O. Random Hypernets in Reliability Analysis of Multilayer Networks. *Electrical Engineering*, 2015, vol. 343, pp. 307-315.
15. Kalney A. M. Optimizaciya razmeshcheniya kontrol'nyh ustrojstv na kanalah v setyah monitoringa [Optimizing the placement of control devices on channels in monitoring networks]. *Problems of computer science*, 2022, no. 4, pp. 28-38.
16. Rodionov A. S., Migov D. A. Obtaining and Using Cumulative Bounds of Network Reliability. *System Reliability*, 2017.
17. Valiant L. The complexity of computing the permanent. *Theoretical Computer Science*, 1979, vol. 8, no. 2, pp. 189-201.

Artyom M. Kalney

Engineer. Laboratory of System Modeling and Optimization, Institute of Computational Mathematics and mathematical geophysics SB RAS (ICMaMG SB RAS, Russia, 630090, Novosibirsk, Academician Lavrentiev Avenue St. 6), e-mail: artemkalnei@yandex.ru, ORCID ID: 0009-0003-4107-7995.

Aleksej S. Rodionov

Dr. of Sci. (Engineering); Head of the Laboratory of System Modeling and Optimization,fic Institute of Computational Mathematics and mathematical geophysics SB RAS (ICMaMG SB RAS, Russia, 630090, Novosibirsk, Academician Lavrentiev Avenue St. 6); Professor of the Department of computer science, Siberian State University of Telecommunications and Information Science (SibSUTIS), phone: +7 383 3326949, e-mail: alrod@sscc.ru, ORCID ID: 0000-0001-8411-4707.

DOI: 10.55648/1998-6920-2024-18-3-57-72 УДК 621.315.61.004.6:621.3.049.77

Отказы интегральных схем, вызванные воздействием эффектов электромиграции и антенны

В. В. Шубин

Сибирский гос. унив. телекоммуникаций и информатики (СибГУТИ)

Аннотация: В статье представлены некоторые проблемы отказов работы интегральных схем (ИС) и их предотвращение на ранних этапах проектирования. Рассмотрены, обобщены и систематизированы вопросы, связанные с отказами ИС, вызванными воздействием эффектов электромиграции и антенны. В статье приведены некоторые примеры, которые могут быть использованы в практической деятельности при разработке ИС для повышения их надёжности с учётом современных тенденций развития в области микроэлектроники.

Ключевые слова: электромиграция, эффект антенны.

Для цитирования: Шубин В. В. Отказы интегральных схем, вызванные воздействием эффектов электромиграции и антенны // Вестник СибГУТИ. 2024. Т. 18, № 3. С. 57–72. https://doi.org/10.55648/1998-6920-2024-18-3-57-72.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Шубин В. В., 2024

Статья поступила в редакцию 10.07.2023;
переработанный вариант – 16.02.2024;
принята к публикации 12.03.2024.

1. Введение

Одной из важнейших задач современной микроэлектроники является повышение надёжности интегральных схем (ИС). Поэтому поиск, анализ и устранение непредсказуемых отказов ИС является весьма важным вопросом исследований. Среди выявленных источников отказов ИС особое место занимают отказы, вызванные электромиграцией и эффектом антенны. Под электромиграцией понимают механизм медленного истощения (вымывания) материала проводника. Эффект антенны представляет собой особый тип пробоя диэлектрика, вызванного накоплением заряда на осаждённых проводниках во время плазмохимического травления, ионной имплантации или процесса удаления фоторезиста.

2. Электромиграция

Электромиграция (ElectroMigration, EM) представляет собой явление медленного износа материала проводника, вызванного протеканием через него электрического тока чрезмерной плотности. Причиной этого явления является захват и постепенное перемещение стационар-

ных атомов проводника носителями электрического тока. Впервые этот феномен был описан В. Б. Фиксом в 1959 г. [1]. Данная работа послужила причиной многочисленных теоретических, экспериментальных и численных исследований данного явления, интерес к которым растёт по мере масштабирования размеров ИС.

2.1. Механизм действия

В чистом алюминии заметный эффект ЕМ наблюдается при плотности тока ~$5 \cdot 10^5$ А/см^2. Данное значение плотности тока может вызывать ЕМ и легко достигается при протекании тока 2.5 мА в проводнике шириной 1 мкм толщиной 5000 Å [**2,3**]. Первая модель эффекта ЕМ, предложенная В. Б. Фиксом и Х. В. Хантингтоном [2], получила название «Баллистическая модель электромиграции». Сущность предложенной модели заключается в рассеянии электронов на дефектах примеси и соответствующей передаче импульса этим дефектам. Данный эффект принято называть «электронным ветром». «Электронный ветер» захватывает и переносит дефекты структуры проводника – примесные атомы, собственные ионы, дефекты междоузлий, вакансий, дислокаций и т.д. В этой модели передача импульса определяется плотностью потока электронов n_e, плотностью дефектов n_d и временем релаксации τ электронов, включая все механизмы рассеяния. Если при столкновении с дефектом электрон передаёт свой импульс $-e\vec{A}\tau$, то электроны за единицу времени передают на один дефект момент силы $n_e e\vec{E}\tau / n_d$. Таким образом, передача импульса за единицу времени, которая является движущей силой «электронного ветра», равна:

$$\vec{F}_{wind} = -\frac{n_e e\tau}{n_d \tau_i}\vec{E} = -\frac{n_e \rho_d}{n_d \rho} e\vec{E}, \qquad (1)$$

где ρ – полное удельное сопротивление металлического соединения и ρ_d – вклад дефектов в полное удельное сопротивление. При комнатной температуре $(\rho_d / n_d)(n_e / \rho) \approx 50$ [4].

Уравнение (1) построено на квантово-механической теории, которая на практике редко применяется из-за сложности вычислений и анализа результатов [5]. Несмотря на свою гомогенную природу, металл алюминия является поликристаллическим материалом. То есть он состоит из отдельных кристаллов, или зёрен (гранул), соединенных встык друг с другом. К сожалению, образование зёрен в проводниках носит непредсказуемый характер и неопределенный размер, и форма этих зёрен может сильно отличаться. Так как носители, образующие ток в проводнике, протекают в основном по границам этих зёрен, легко объяснить локальное отличие плотности тока на отдельных участках этого проводника (рис. 1).

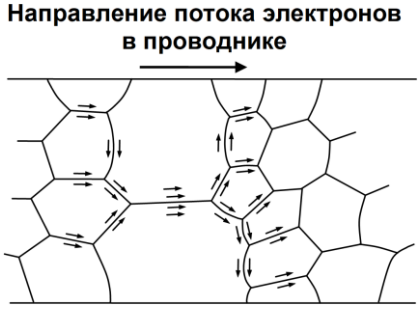


Рис. 1. Демонстрация локального повышения плотности тока в металлическом проводнике

ЕМ заставляет атомы металла постепенно перемещаться вдоль границ зёрен, образуя пустоты между смежными гранулами (рис. 2). Поэтому площадь сечения проводника уменьшается, а плотность тока на оставшемся проводнике увеличивается. Возникающие пустоты постепенно объединяются до тех пор, пока они не сформируют полный разрыв проводника.

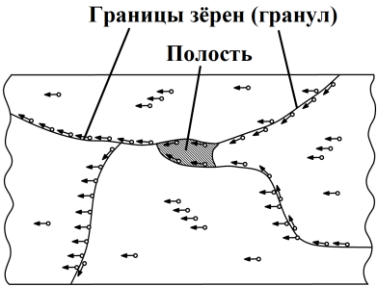


Рис. 2. Возникновение пустот на зауженных фрагментах гранул

Металл, перемещённый из пустот, образует небольшие выпуклости, называемые «бугорками», или вытесняется в заострённые дендриты (пикообразные отростки), которые могут закорачивать соседние проводники между собой. Примеры образования пустот в слое алюминиевой металлизации при передаче сигнала из слоя металл-2 (M2) в слой металл-1 (M1) представлены на рис. 3 [5].

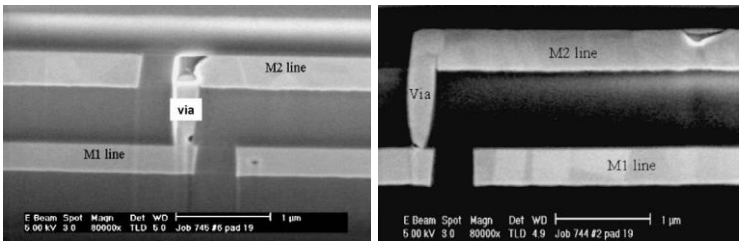


Рис. 3. Примеры образования пустот в слое алюминиевой металлизации

Современные технологические процессы изготовления ИС для защиты от отказов ЕМ предлагают дополнение к алюминиевой металлизации в виде плёнок тугоплавких металлов – W или Ti. Тугоплавкие металлы обладают более высоким удельным сопротивлением, вследствие чего большая часть тока протекает через алюминий. Кроме того, тугоплавкие металлы существенно меньше восприимчивы к эффекту ЕМ. Поэтому, если эффект ЕМ в конце концов вымывает алюминий из проводника, то плёнка тугоплавкого металла продолжает проводить ток и функционально предотвращает фатальный отказ ИС в виде полного обрыва этого проводника (рис. 4).

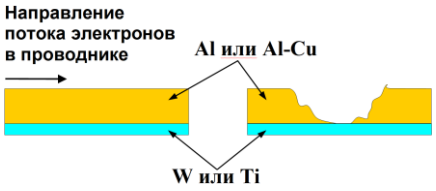


Рис. 4. Сохранение функциональной работоспособности схемы
после вымывания Al из алюминиевого проводника, защищённого тугоплавким металлом

Однако формирование пустот в Al служит причиной постепенного и непредсказуемого роста сопротивления проводников узлов, соединяющих компоненты в требуемую схему. При этом следует учитывать, что даже сохранение функциональной работоспособности ИС не означает сохранение полноценной её работы. Изменение сопротивления нарушает динамические характеристики схемы или приводит к рассогласованию сигналов, требующих прецизионного согласования. Поэтому изготовитель ИС для операций металлизации устанавливает ограничения на величину поперечного сечения Al части проводника и величину тока, при котором это сечение может безопасно проводить этот ток, независимо от наличия или отсутствия тугоплавкого защитного металла. Ещё более угрожающе выглядит металлический про-

водник Al, перемещаемый из пустот, который формирует дендриты, закорачивающие соседние проводники между собой. На рис. 5 [5] приведён пример образования дендритов.

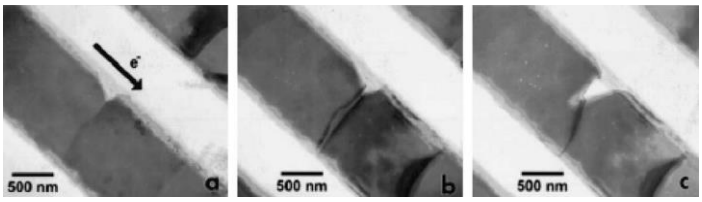


Рис. 5. Развитие дендритов (заострённых пикообразных отростков)

Кроме того, защита тугоплавкими металлами часто используется для предотвращения отказов, вызванных ЕМ между различными уровнями металлизации в контактных окнах (КО) и переходных отверстиях (ПО). Наличие тугоплавких металлов по границам КО и ПО гарантирует сохранность электрического сигнала в узле, даже если алюминиевая металлизация в них становится жертвой пустот, вызванных ЕМ (рис. 6).

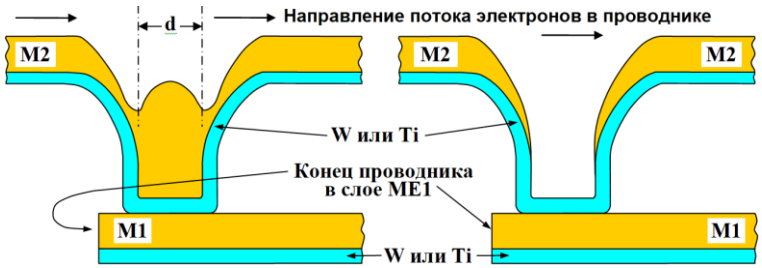


Рис. 6. Схема образования пустот в переходных отверстиях M2–M1 (ПО 2-1)

В случае на рис. 6 сигнал проводника в слое M2 относительно ПО – проходной. Поэтому образование пустот может привести к полному вытеснению Al из КО и ПО на их границу в сторону направления потока носителей. Однако во многих случаях КО и ПО являются концом проводника, переходящего на начало проводника более низкого уровня. В этом случае в КО и ПО вытеснение алюминия не происходит, так как возможность вытеснения закрывается тонкой плёнкой тугоплавкого металла, который полностью покрывает стенки структуры отверстий, как показано на рис. 7 [5].

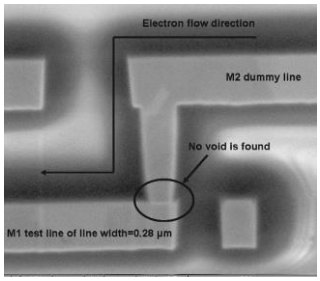


Рис. 7. Пример сохранения переходного отверстия в соединении M2 с M1

Последние исследования подтверждают более ранние, которые установили, что в ПО наиболее уязвимым местом с точки зрения образования пустот вследствие ЕМ является собственно соединение между алюминием и защитным тугоплавким металлом. Установлено, что противодействие ЕМ в такой неоднородной системе, как ПО, действительно зависит от направления потока носителей. При направлении потока носителей, как показано на рис. 8 [5], образование пустот происходит в нижнем слое металлизации – M1. При обратном направлении потока носителей вымывание металла происходит в верхнем слое металли-

зации – M2, на границе ПО в направлении потока носителей [6]. Таким образом, перед разработчиками технологического процесса стоит непростая задача: определение и сертификация реального качества КО и ПО, подтверждённого экспериментальными измерениями и испытаниями [2].

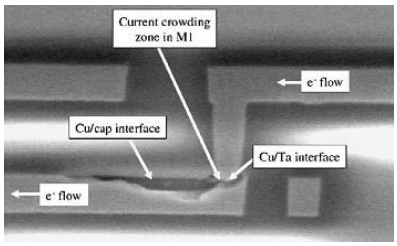


Рис. 8. Образование пустот в нижнем слое металлизации на границе ПО между слоями металлизации M2 и M1 при направлении потока носителей из верхнего слоя металлизации M2 в нижний M1

2.2. Методы противодействия отказам ИС, вызванным эффектом ЕМ

Первая линия защиты от воздействия ЕМ заключается в совершенствовании технологического процесса изготовления ИС. Современные технологические процессы для усиления противодействия ЕМ обычно применяют легирование Al-металлизации медью от 0.5 до 4% [7]. Медь скапливается на границах зёрен, где она препятствует возникновению пустот за счёт увеличения энергии активации, необходимой для перемещения атомов из атомной решетки. Алюминий, легированный медью, показывает ~5–10-кратное улучшение способности его обработки по сравнению с чистым алюминием [8]. Противодействие проводников ЕМ также может быть повышено путём использования защитных покрытий, полученных под давлением, за счёт уплотнения металла, в котором формирование пустот существенно затрудняется [9].

Чистая медь более устойчива к воздействию эффекта ЕМ по сравнению с чистым алюминием или даже алюминием, легированным медью. Заметное увеличение времени жизни проводника, подверженного действию эффекта ЕМ, зависит от происхождения межуровневого диэлектрика, механизма пассивации металла проводников и от условий испытаний. Типовым значением продления срока жизни алюминия является его увеличение в 40–100 раз [10]. Система металлизации из образованной толстым слоем меди фактически не подвержена эффекту ЕМ. При этом «дамасский» процесс получения меди и даже двойной «дамасский» повышают устойчивость системы металлизации, но не могут гарантировать полной защиты от эффекта ЕМ.

Приведённые технологические методы позволяют минимизировать эффект ЕМ, но всё равно существует некая предельная плотность тока, превышение которой при соответствующих условиях приводит к возможности отказа ИС вследствие нарушений системы металлизации. Поэтому правила проектирования любого технологического процесса изготовления ИС должны обязательно устанавливать величину удельной предельно допустимой плотности тока на единицу ширины проводника. Типовое значение удельной плотности тока металлического проводника J_S обычно составляет 1 мА/мкм для проводников, пересекающих ступеньку оксида или поликремния, и 2 мА/мкм для проводников, проходящих по гладкой поверхности полевого оксида. Эти значения зависят от состава слоя металла, его толщины и предельно допустимой величины рабочей температуры. Выражение (2) описывает зависимость плотности тока J от допустимой ширины проводника w:

$$J = J_S \cdot w, \tag{2}$$

Рассмотрим пример, в котором проводник должен проводить 50 мА при типовых значениях J_S. Согласно выражению (2), можно вычислить минимальную ширину проводника, допускаемую антиэлектромиграционными ограничениями правил проектирования: $w = J / J_S$. Таким образом, для проводников, проходящих по полевому оксиду, минимальная ширина должна составлять: w = 50/2 = 25 мкм, и для проводников, проходящих по ступеньке оксида и (или) поликремния: w = 50/1 = 50 мкм. Кроме того, правила проектирования технологического процесса должны запрещать резкое расширение проводника на оксидной или поликремниевой ступеньке (рис. 9A) и, так как ток должен втекать из узкого в более широкий проводник постепенно, точно и однозначно представлять требования формирования схемы постепенного расширения проводника. Обычно расширение проводника должно распространяться в обе стороны относительно ступеньки на расстояние по крайней мере удвоенной наибольшей ширины, как показано на рис. 9B–E.

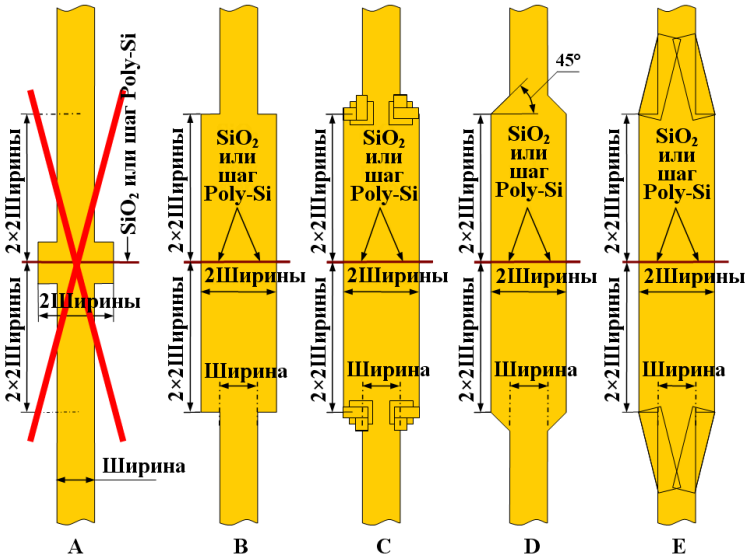


Рис. 9. Пример топологии различных схем расширения проводника на ступеньках оксида или поликремния для подавления эффекта ЕМ: A – недопустимое расширение проводника; B – допустимое расширение проводника; C – улучшенное допустимое расширение проводника; D – наиболее распространенное расширение проводника; E – идеальное расширение проводника

Большинство правил проектирования определяют максимальный ток, допустимый для протекания через КО и ПО, и устанавливают величину тока, который может протекать через проводник такой же ширины, что и размер КО и ПО. Следуя этому принципу, ПО шириной 1 мкм могут проводить ток такой же силы, что и проводник шириной 1 мкм или, как было указано ранее, 2.5 мкм. Заметим, что значения токов, приведённых в данном примере, являются типовыми и не обязательно должны распространяться на любой технологический процесс.

Чрезмерный ток также может служить причиной перегрева и последующих отказов проволочек, соединяющих контактные площадки кристалла с траверсами (рамкой) корпуса. Если предполагаемый ток превышает пределы, установленные правилами проектирования, то проект требует применения проволочек большего диаметра или параллельного мультиплицирования несколькими проволочками при сварке одного траверса корпуса с одной контактной площадкой кристалла [2]. В общем случае плотность тока J определяется как частное от деления силы тока I и площади поперечного сечения проводника S:

$$J = \frac{I}{S}. \qquad (3)$$

Конструкция топологии ИС оказывает существенное влияние на проявление эффекта ЕМ по всему кристаллу. Особое внимание разработчиков топологии ИС заслуживает проблема трассировки сигналов проводниками с изменением направления не только в вертикальном, но и горизонтальном направлении. Очевидно, что в топологии любое изменение направления проводников с поворотом на 90° на коротком участке поворота приводит к локальному увеличению плотности протекающего тока, как показано на рис. 10 [11].

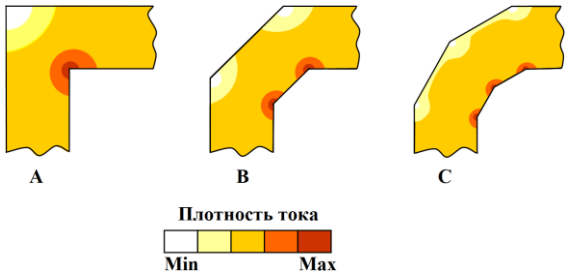


Рис. 10. Визуализация плотности тока при различных изгибах внутренних углов проводников ИС: A – изгиб угла 90°; B – изгиб угла 135°; C – изгиб угла 150°

Из рис. 10 видно, что самая высокая плотность тока и, следовательно, вероятность разрушения проводника воздействием эффекта ЕМ возникает при использовании топологии с углом поворота 90° (рис. 10A). Поэтому для увеличения срока службы слоя металлизированной разводки разработчик топологии должен более внимательно относиться к выбору стратегии трассировки проводников при изменении направления распространения сигналов.

3. Эффект антенны

Одной из важнейших операций современных технологических процессов изготовления ИС в области «глубокого субмикрона» (Deep SubMicron, DSM) является операция сухого травления. Сухое травление, или плазмохимическое травление (ПХТ), представляет собой операцию удаления плёнки материала путем воздействия на материал бомбардировкой ионов, которые вытесняют фрагменты материала с открытой поверхности. Чаще всего ПХТ использует плазму реакционноспособных газов, таких как фторуглерод, кислород, хлор, трихлорид бора, иногда с добавлением азота, аргона и гелия. Процесс ПХТ является анизотропным и мало влияет на изменение геометрии рисунка материала [2].

Установлено, что в процессе планарного ПХТ на поверхности пластины осаждаются заряды. Поэтому любой обнаженный проводник может аккумулировать электрический заряд. Этот заряд при достижении определённого значения может повредить тонкий диэлектрик, расположенный под проводником уже на этапе изготовления. Этот механизм отказа называется механизмом разрушения, вызванного процессом ПХТ (Process plasma-induced damage). В профессиональном сообществе разработчиков ИС этот эффект принято называть эффектом антенны (Antenna Effect, AE). Разрушения, вызванные этим эффектом, могут увеличить утечку через оксид затвора MOS-транзистора в статическом режиме, изменить его пороговое напряжение или сократить срок службы транзистора [12].

3.1. Механизм действия

Точный источник электрических зарядов, отвечающий за возникновение эффекта антенны, окончательно не установлен и остаётся предметом дискуссии. Плазма сама по себе состоит из равного количества положительно и отрицательно заряженных частиц. Однако различные механизмы могут вызывать локальные флуктуации плотности заряда внутри плазмы.

Чаще всего причины этих флуктуаций объясняются:

• отклонениями однородности конструкции реактора;

• АС-активацией плазмы, или эффектом «электронного затемнения» (Electron shading), в котором смежные геометрические фигуры блокируют изотропный поток электронов в большей степени, чем они блокируют анизотропный поток ионов.

Тем не менее, независимо от того, какой именно механизм вовлечён в процесс аккумуляции зарядов на проводники, опыт показывает, что как ПХТ слоёв проводников, так и последующее озоление фоторезиста (ashing – сжигание, удаление фоторезиста в кислородной плазме) могут служить причинами разрушений, вызываемых ПХТ (plasma-induced damages). Рис. 11 позволяет понять механизм накопления заряда на поверхности проводника.

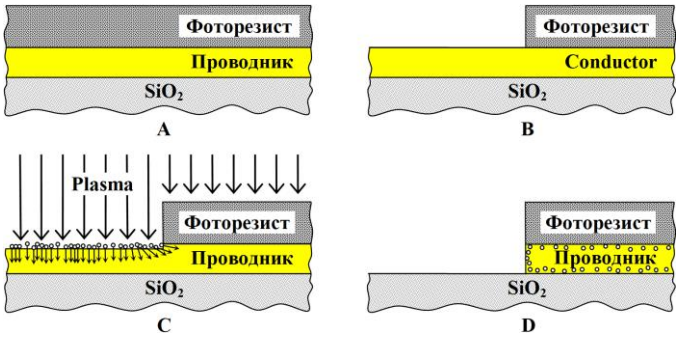


Рис. 11. Механизм накопления заряда на поверхности проводника: A – Нанесение защитного слоя; B – Формирование рисунка в защитном слое; C – Начало ПХТ. Концентрация зарядов на открытой поверхности проводника; D – Перераспределение зарядов на сформированные проводники

Так как в современных технологических процессах изготовления ИС в травление всех проводящих слоёв вовлечена операция ПХТ, то воздействие эффекта антенны должно быть оценено для каждого из этих слоёв и соответствующие ограничения внесены в правила проектирования.

Чаще всего наиболее тонким диэлектриком, разделяющим проводящие слои в современных CMOS-процессах изготовления ИС, является оксид поликремниевого затвора MOS-транзистора (MOS transistor poly gate oxide). Поэтому, как правило, при формировании правил антенны основное внимание уделяется ограничениям создания рисунка топологии первого уровня поликремния. Так как первый уровень поликремния является проводящим слоем, то эффект антенны может служить источником разрушения затворного диэлектрика уже на этапе формирования рисунка топологии затвора и удаления фоторезиста методом озоления.

Во время начальных стадий травления поликремниевого затвора вся поверхность пластины (wafer) закрыта непрерывной плёнкой поликремния. Заряд попадает на эту плёнку через все открытые поверхности в фоторезисте (рис. 11B). Флуктуации заряда, отвечающие за возникновение эффекта антенны, после окончания ПХТ перераспределяются на изолированные участки вдоль направлений поперечных сечений поликремниевого примитива. В процессе ПХТ отдельные участки поликремния частично открываются под воздействием плазмы (рис. 11C). Каждая геометрическая фигура теперь приобретает заряд вокруг своей периферии, где поликремний открывается воздействию плазмы. Этот заряд инжектируется в тонкий затворный окисел (рис. 11D). Следовательно, уязвимость данной геометрической фигуры к эффекту антенны зависит от отношения общего периметра примитива затвора к активной площади на тонком оксиде (площади канала). Это соотношение называется периферическим коэффициентом антенны (Peripheral Antenna Ratio, PAR). Чем больше величина PAR, тем больше риск разрушения, вызванного эффектом антенны. Поэтому большинство технологических процессов обязательно регламентируют значение PAR для поликремния. Стандартное значение PAR для CMOS-процесса для 180 нм обычно составляет 100 мкм$^{-1}$.

Во время завершающей стадии озоления фоторезиста вся поверхность поликремния становится открытой для плазмы. Теперь каждая геометрическая фигура аккумулирует заряд на всю поверхность вскрытых поликремниевых затворов. Этот заряд воздействует на тонкий затворный оксид. Так как заряд распределяется более или менее равномерно по всей поверхности затвора, уязвимость подзатворного оксида к эффекту антенны будет определяться отношением площади каждого поликремниевого примитива, образующего затвор, к площади его тонкого подзатворного окисла, которая формирует область канала. Это соотношение называется коэффициентом антенны площадей (Areal Antenna Ratio, AAR). Чем больше величина AAR, тем больше риск разрушений, вызываемых ПХТ. Большинство технологических процессов для поликремния определяют величину AAR. Обычно фабрики-изготовители для CMOS-процесса для 180 нм устанавливают значение AAR, равное 500 [2].

Так как размеры затворов MOS-транзисторов ограничены другими правилами проектирования и, как правило, невелики, то отказы, вызванные эффектом антенны непосредственно в процессе формирования поликремния, возникают крайне редко. Бóльшую опасность для затворов MOS-транзисторов представляет накопление заряда на проводниках, соединяющих затворы с другими элементами схемы (рис. 12), и его передача на затвор, так как длина этих проводников неизвестна и может превышать линейные размеры кристалла.

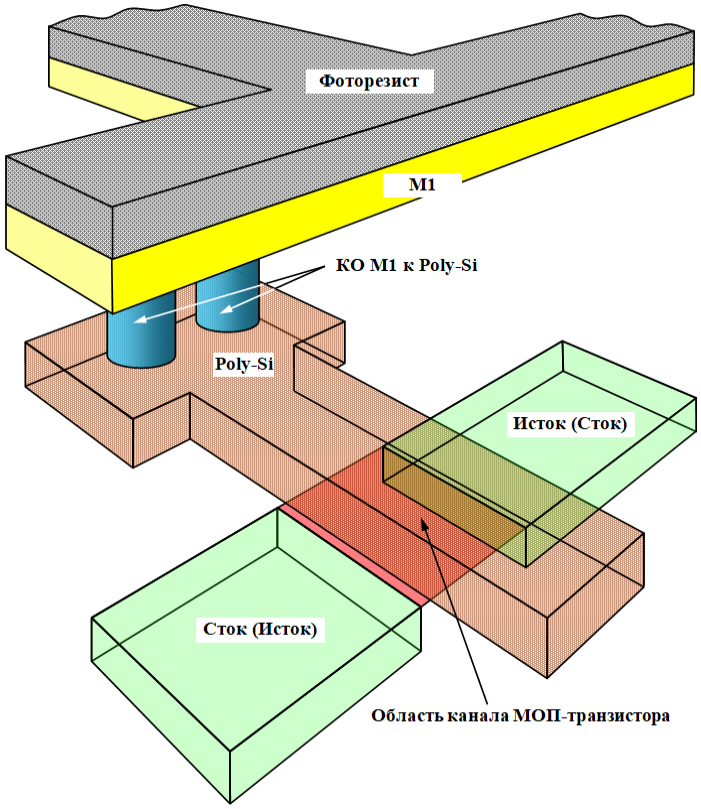


Рис. 12. Схема MOS-транзистора и формирования узла Metal-1/Poli-Si

Любой проводящий слой трассировки ИС подвержен воздействию эффекта антенны во время плазмохимического травления и озоления, так что для каждого из них определяются свои собственные коэффициенты антенны: периферический (PAR) и соотношения площадей (AAR). В качестве примера рассмотрим случай формирования низкоомного слоя второго уровня трассировки M2. В конце процесса травления отдельные геометрические фигуры в слое M2 становятся отделёнными одна от другой. Однако эти геометрические фигуры могут быть соединены вместе через нижние проводящие слои, тем самым образуя единый электрический узел. В этом случае эффект антенны не может оцениваться на основе последовательного контроля одной геометрической фигуры за другой. Вместо этого необходимо контролировать набор электрически соединённых геометрических фигур, образующих отдельные уз-

лы. В процессе травления слоя M2 каждый узел собирает заряд пропорционально периферийной поверхности слоя, открытого для плазмы, и инжектирует этот заряд вплоть до геометрических примитивов активных поликремниевых затворов, формирующих часть узла. Поэтому PAR слоя M2 узла равняется общему периметру узла в слое M2, поделенного на активную площадь геометрического примитива поликремниевого затвора в узле (общая площадь поликремниевых затворов, подсоединённых к узлу). Аналогично эволюция отказов, вызванных процессом озоления, зависит от величины AAR слоя M2 [13, 14].

Все сигналы, поступающие на затворы MOS-транзисторов, передаются по проводникам на разных уровнях металлизации. Поэтому в правилах проектирования устанавливаются ограничения, связанные с поражением затворного оксида, вызванного эффектом антенны для каждого уровня металлизации. В правилах проектирования правила антенны (Antenna Rules) описываются ограничением величины AR_n. Правило антенны для *n*-го слоя металлизации AR_n показывает, во сколько раз открытая текущим процессом травления площадь поверхности узла проводящего слоя, подсоединённого к затвору, может быть больше площади тонкого окисла затвора. Наличие соединения расчётного узла с P-(N-)-переходом (диодом) снимает ограничение со всех последующих слоёв, соединённых с этим узлом.

$$AR_n = S_A / S_{TG} \ , \tag{4}$$

где S_A – площадь антенны (площадь боковой поверхности узла, подсоединенного к затвору), S_{TG} – топологическая площадь канала.

$$S_A = P_1 \cdot H_1 + P_2 \cdot H_2 + ... + P_n \cdot H_n \, , \tag{5}$$

где P – периметр проводящего слоя антенны, H – толщина проводящего слоя антенны, n – номер проводящего слоя.

$$S_{TG} = W_i \cdot L_i \, , \tag{6}$$

где W – топологическая ширина канала, L – топологическая длина канала, i – порядковый номер узла.

Несмотря на то, что на исследования взаимосвязи между влиянием эффекта антенны на отказы, связанные с разрушением затворного оксида, было затрачено множество усилий, до сих пор полного понимания этой проблемы не существует. Отдельные исследования находят доказательства того, что PMOS затворные оксиды значительно более чувствительны к механизму разрушения, вызванного технологическим процессом ПХТ (Process plasma-induced damage), чем NMOS затворные оксиды. Поэтому некоторые технологические процессы определяют отдельно антенные коэффициенты для каждого типа оксида. Другие исследования показали, В других исследованиях было показано, что вероятность разрушений изолирующего оксида, вызванных процессом ПХТ, может быть существенно уменьшена ограничением тока, протекающего через какую-либо область затворного оксида [15]. Результаты этих исследований открывают дополнительные возможности поиска конструктивно-технологических способов подавления эффекта антенны.

Строго говоря, любой оксид может быть подвержен воздействию эффекта антенны, в том числе полевой, межслойный, LOCOS, STI и т.д. Другое дело, что значения пробивных напряжений этих оксидов многократно превышают величины возможных накоплений зарядов, способных привести к разрушению этих оксидов. Поэтому ограничения, необходимые для защиты этих оксидов от воздействия эффекта антенны, в настоящее время не регламентируются.

В современных технологических процессах изготовления ИС существуют рабочие приборы, которые требуют формирования тонкого оксида не менее уязвимого к воздействию эффекта антенны, чем затворный оксид MOS транзисторов, например, конденсаторы «металл – металл» (Metal-insulator-Metal, MiM). В данном конденсаторе между любыми двумя слоями металлизации может быть внедрен дополнительный слой металла верхней обкладки CTM (Capacitor Top Metal). Структура такого конденсатора показана на рис. 13.

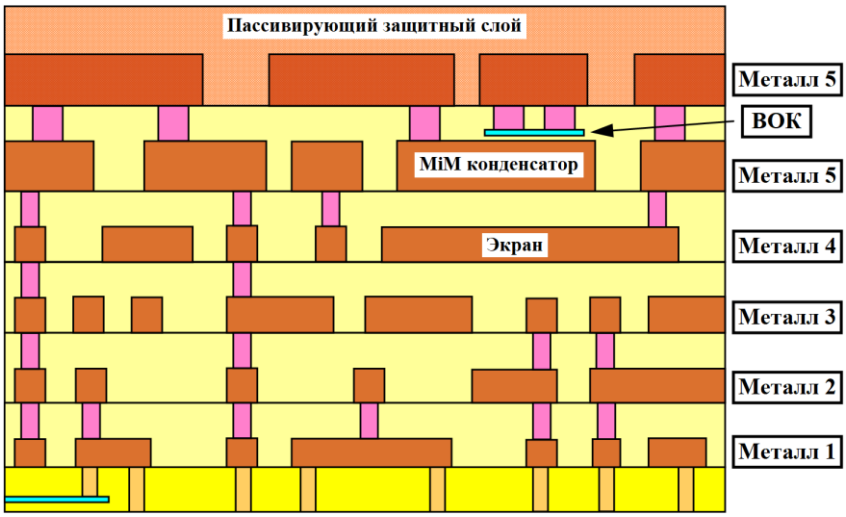


Рис. 13. Эскиз структуры MiM-конденсатора между шестым и пятым уровнями металлизации

3.2. Методы противодействия отказам ИС, вызванным воздействием эффекта антенны

Ограничения правил проектирования, установленные для противодействия эффекту антенны, регламентируются правилами антенны. Любая фигура топологии или узел проводящего слоя контролируются и в случае несоответствия требованиям должны быть переработаны. Конкретные схемо-топологические методы, которые используются, зависят от конкретного слоя. В настоящее время для уменьшения воздействия эффекта антенны на разрушение подзатворного диэлектрика во время формирования рисунка поликремния используют два способа [16]:

1. Уменьшение величины отношения площади затвора к площади канала MOS-транзистора до значений AAR, разрешённых правилами проектирования, за счёт увеличения длины канала, ширины канала или длины и ширины канала одновременно (рис. 14A). Такие изменения могут привести к уменьшению крутизны MOS-транзисторов и увеличению паразитной ёмкости узла, которые требуют обязательной оценки их влияния на функционирование ИС и точно снижают её надёжность и долговечность.

2. Уменьшение величины отношения площади затвора к площади канала MOS-транзистора до значений AAR, разрешённых правилами проектирования, за счёт уменьшения площади затвора путём введения перехода в слой M1 (рис. 14B). Такое изменение может привести к увеличению паразитного сопротивления узла для поступающего полезного сигнала, поэтому необходима обязательная оценка их влияния на функционирование ИС. Кроме того, разработчик топологии должен учитывать тот факт, что рост количества связей в ИС также снижает её надёжность и долговечность.

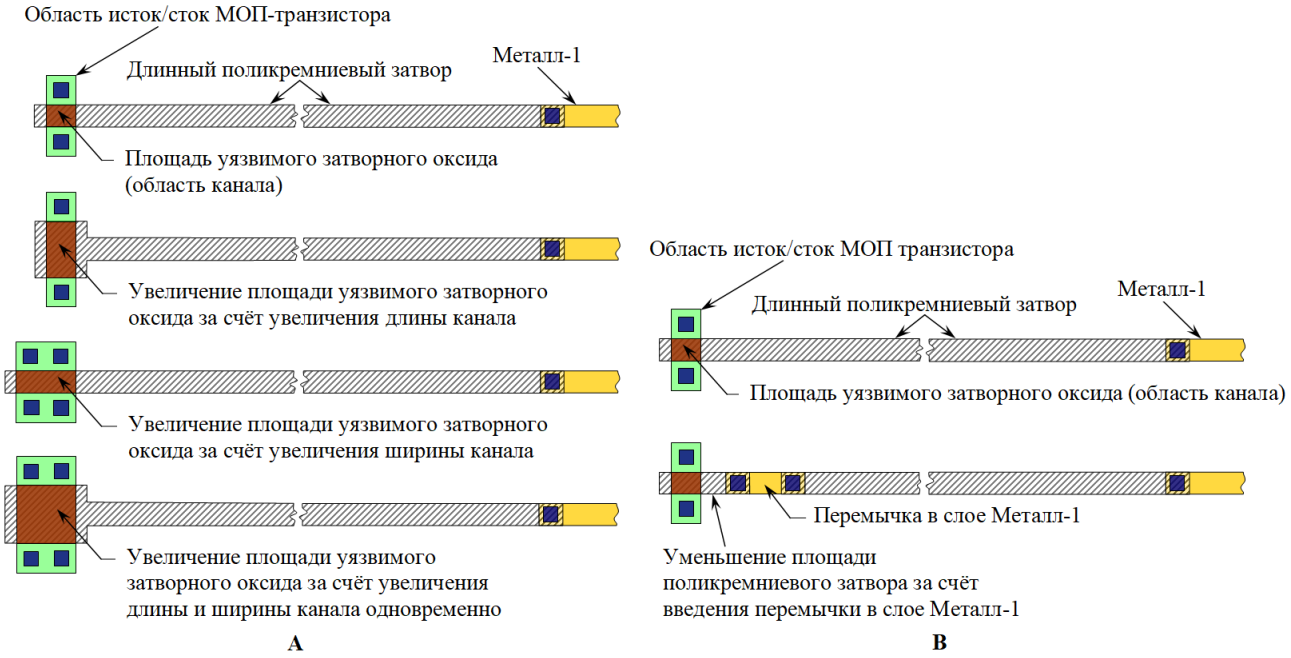


Рис. 14. Механизмы подавления эффекта антенны во время формирования рисунка затворного поликремния: A – Уменьшение величины отношения площади затвора к площади канала MOS-транзистора за счёт увеличения длины канала, ширины канала или длины и ширины канала одновременно; B – Уменьшение величины отношения площади затвора к площади канала MOS-транзистора за счёт уменьшения площади затвора путём введения перехода в слой M1

В любом случае большинство внутренних сигналов ИС формируются на стоках/истоках MOS-транзисторов. Для технологических процессов, в которых применяются затворные оксиды толщиной более 400 Å, исток/стоковые переходы MOS-транзисторов, как правило, переходят в состояние лавинного пробоя раньше, чем могут разрушиться затворные оксиды. В этом случае исток/стоковые области выполняют функцию сточного диода антенны [15]. Такие узлы при вычислении коэффициентов антенны в целом могут быть игнорированы. Однако если узел металла определяется как узел с чрезмерным значением коэффициента антенны и в нем отсутствует соединение с исток/стоковыми областями, то проблема может быть устранена одним из двух способов: размещением перемычки (jumper) на более высокий уровень металлизации или подсоединением структуры, называемой сточный диод антенны (leaker antenna diode). В зависимости от конкретных топологических возможностей сточные диоды могут быть двух типов: NSD/P-epi и PSD/N-well (рис. 15). Вследствие более высокой подвижности электронов по сравнению с дырками для тонкооксидных технологических процессов более предпочтительным является использование сточного диода NSD/P-epi. В сущности, структура представляет собой P-N-переход, образующий диод, чей анод соединён с узлом в первом уровне металлизации, а катод – с подложкой. Если напряжение на узле падает ниже потенциала подложки, то сточный диод антенны NSD/P-epi смещается в прямом направлении и потенциал узла шунтируется на подложку. Если напряжение на узле поднимается выше потенциала подложки, то сточный диод антенны NSD/P-epi переходит в состояние лавинного пробоя раньше, чем наступают разрушения затворного оксида [2].

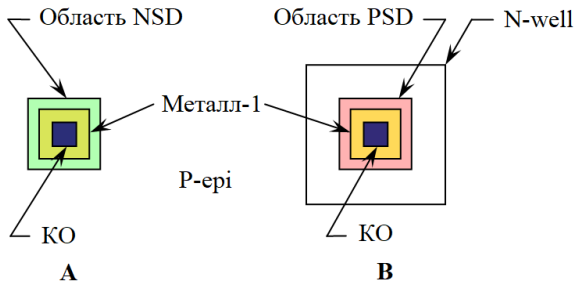


Рис. 15. Сточные диоды антенны: A – сточный диод NSD/P-epi; B – сточный диод PSD/N-well

Более проблематичным представляется использование сточных диодов антенны в технологических процессах, использующих более тонкие оксиды. Лавинный пробой перехода NSD/P-epi не может полноценно защитить затворный оксид тоньше 400 Å. Исследования и последующие испытания показали, что узлы в технологических процессах, использующих более тонкие оксиды, могут быть защищены совместным применением сточных диодов антенны NSD/P-epi и PSD/N-well, как показано на рис. 16 [12]. Такая защита аналогична элементу входной защиты от пробоя ESD (ElectroStatic Discharge) [17]. Переход NSD/P-epi сточного диода смещается в прямом направлении, если потенциал узла падает ниже потенциала подложки P-epi, а переход PSD/N-well сточного диода – если потенциал узла возрастает выше потенциала N-well.

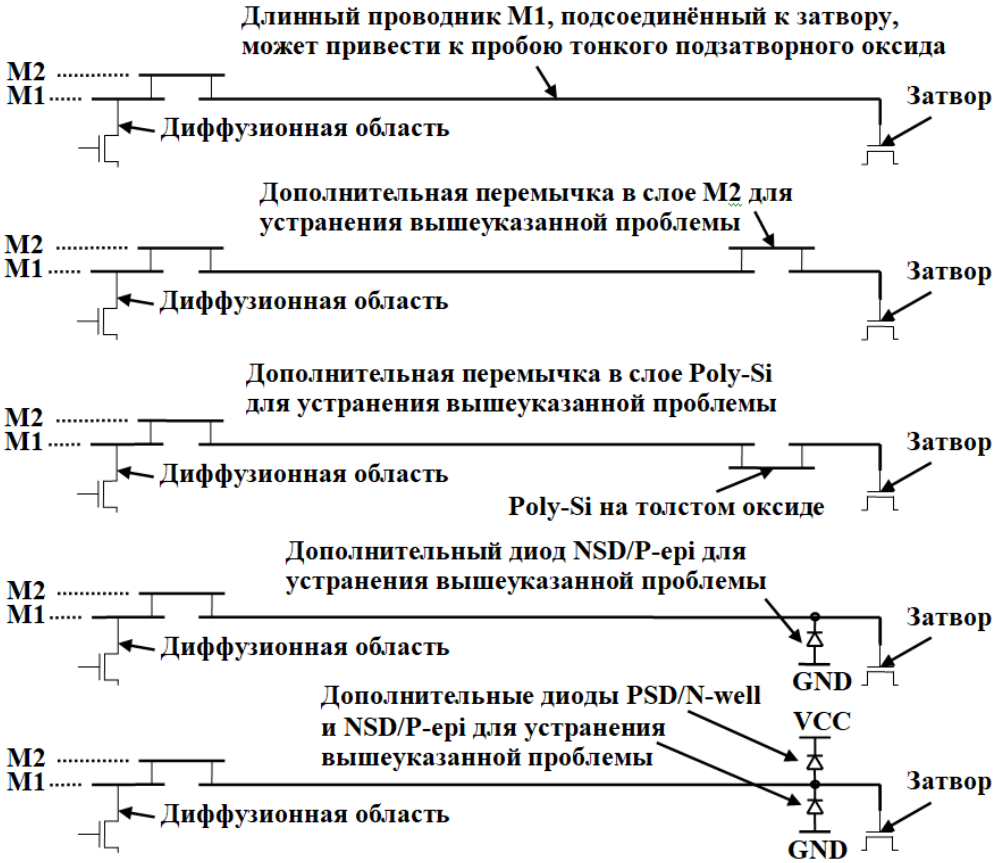


Рис. 16. Схематичное представление нарушения правила антенны и способов его устранения

Переходы сточных диодов антенны NSD/P-epi и PSD/N-well в рабочем режиме ИС смещены в обратном направлении и поэтому не нарушают нормальное функционирование схемы. Однако разработчик должен учитывать, что наличие этих диодов вносит дополнительную паразитную ёмкость P-N-переходов, подсоединённую к узлу, которая изменяет динамические характеристики переходного процесса переключения узла.

Кроме того, во время нормальной работы схемы обратно смещённый переход N-well/P-epi также предотвращает ток утечки через эту структуру. Во время операции ПХТ пламя плазменного процесса освещает поверхность пластины. Это освещение поддерживает фотогенерацию носителей зарядов в области обеднения перехода N-well/P-epi, вызывая его утечку. Данная утечка способствует рассасыванию инжектированного заряда в N-well. Для того чтобы сточный диод N-well/P-epi функционировал должным образом, по крайней мере часть его периферии не должна перекрываться металлом на расстоянии 5–10 мкм от границы рисунка N-well. При каждом введении сточных диодов антенны обязательным требованием является уведомление схемотехника о конструктивном изменении схемы, чтобы он имел возможность оценить влияние этого изменения на функционирование схемы. В большинстве случаев сточные диоды антенны не оказывают влияния на работу схемы, однако данное утверждение, как правило, не распространяется на аналоговые схемы [2].

Для схем, требующих хорошего согласования, таких как аналоговые схемы или схемы, содержащие элементы памяти, для защиты от нарушений затворного оксида, которые могут внести рассогласование, затворы транзисторов должны быть защищены непосредственным соединением к диффузионным областям с коротким сегментом в слое M1.

В заключение заметим, что описанная проблема влияния эффекта антенны на оксид поликремниевого затвора, как и методы противодействия этому влиянию, в равной степени относится к изолирующему оксиду рабочих MiM-конденсаторов, встроенных в многослойную систему трассировки (рис. 13).

Данная статья является третьей частью продолжения цикла материалов по теме «Отказы ИС». Ранее первая часть задуманного цикла «Отказы интегральных схем, вызванные пробоем диэлектрика» была опубликована в журнале «Вестник СибГУТИ», Том 17, № 2, 2023 г., вторая часть «Отказы интегральных схем, вызванные эффектами распространения зарядов и образования паразитных каналов» – в журнале «Электронная техника. Серия 3. Микроэлектроника», выпуск 1 (193).

Литература

1. *Fiks V. B.* On the mechanism of the mobility of ions in metals // Soviet Physics: Solid State, 1959. V. 1, № 1. P. 14–28.
2. *Hastings Alan.* The Art of ANALOG LAYOUT. New Jersey. Pearson Prentice Hall, 2006. 648 p.
3. *Huntington H. B., Grone A. R.* Current-induced marker motion in gold wires // Journal of Physics and Chemistry of Solids. 1961. V. 20. P. 76–87.
4. *Verbruggen A. H.* Fundamental questions in the theory of electromigration // IBM Journal of Research and Development. 1988. № 32. P. 93–98.
5. *Tan C. M., Arijit R.* Electromigration in ULSI Interconnection // Materials Science and Engineering. Elsevier. 2007. Is. 58. P. 1–75.
6. *Tao J., Young K. K., Cheung N. W., Hu C.* Comparison of Electromigration Reliability of Tungsten and Aluminum Vias Under DC and Time-Varying Current Stressing // Proc. International Reliability Physics Symposium, 1992. V. 14, № 4. P. 338–343.
7. *Ames I., d'Heurle F. M., Horstmann R. E.* Reduction of Electromigration in Aluminium Film by Copper Doping // IBM Journal of Research and Development. 1970. V.14, № 4. P. 461–463.
8. *Iyer S. S., Ting Ch. Y.* Electromigration Study of Al-Cu/Ti/Al-Cu System // Proc. International Reliability Physics Symposium, 1984. P. 273–278.
9. *Lloyd J. R., Smith P. M.* The Effect of Passivation on the Electromigration Lifetime of Al/Cu Thin Film Conductors // Journal of Vacuum Science Technology A. 1983. V. 1, № 2. P. 455–458.

10. *Chen J. Z., Amerasekera A., Duvvury Ch.* Design Methodology for Optimizing Gate Driven ESD Protection Circuits in Submicron CMOS Processes // Proc. EOS/ESD Symposium, 1997. P. 1–10.
11. *Lienig J, Thiele M.* Fundamentals of Electromigration-Aware Integrated Circuit Design. SpringerLink, 2018. 159 p.
12. *Weste N. H. W., Harris D. M.* CMOS VLSI Design: A Circuits and Systems Perspective. 4nd Ed. Boston: Addison-Wesley, Pearson Education, Inc., 2011. 838 p.
13. *Gabriel C. T.* Gate Oxide Damage: A Brief History and a Look Ahead // Proc. 6th International Symposium on Plasma Process-Induced Damage, 2001. P. 20–24.
14. *Watanabe T., Yoshida Y.* Dielectric Break-down of Gate Insulator Due to Reactive Ion Etching // Solid State Technology. 1984. V. 27, № 44. P. 263–266.
15. *Mocuta A. C., Hook T. B., Chou A. I., Wagner T., Stamper A. K., Khare M., Gambino J. P.* Plasma Charging Damage in SOI Technology // Proc. 6th International Symposium on Plasma Process-Induced Damage, 2001. P. 104–107.
16. *Simon P., Luchies J.-M., Maly W.* Antenna Ratio Definition for VLSI Circuits // Proc. 4th International Symposium on Plasma Process-Induced Damage, 1999. P. 16–20.
17. *Baker R. Jacob*. CMOS: Circuit Design, Layout, and Simulation. New Jersey: John Wiley & Sons Inc., 2011. 1208 p.

Шубин Владимир Владимирович
к.т.н., доцент кафедры технической электроники СибГУТИ;
начальник отдела по разработке аналоговых ИМС АО «НЗПП Восток» (630082, Новосибирск, ул. Дачная, 60), e-mail: shubin@nzpp.ru, ORCID ID: 0000-0002-2974-0497.

Автор прочитал и одобрил окончательный вариант рукописи.
Автор заявляет об отсутствии конфликта интересов.

Integrated Circuits Failures Caused by Electromigration and Antenna Effects

Vladimir V. Shubin

Siberian State University of Telecommunications and Information Science (SibSUTIS)

Abstract: The paper describes some problems of operation failures of integrated circuits (ICs) and corresponding preventative measures at the early stages of design process. The issues related to the problems of IC failures caused by electromigration and antenna effect are considered, generalized and systematized. Some examples that can be used in practice when developing ICs to improve their reliability taking into account current trends in the microelectronics fields are presented.

Keywords: electromigration, antenna effect.

For citation: Shubin V. V. Integrated Circuits Failures Caused by Electromigration and Antenna Effects (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 3, pp. 57-72. https://doi.org/10.55648/1998-6920-2024-18-3-57-72.

Content is available under the license Creative Commons Attribution 4.0 License

© Shubin V. V., 2024

The article was submitted: 10.07.2023;
revised version: 16.02.2024;
accepted for publication 12.03.2024.

References

1. Fiks V. B. On the mechanism of the mobility of ions in metals. Soviet Physics: Solid State, 1959, vol. 1, no. 1, pp. 14-28.
2. Hastings Alan. The Art of ANALOG LAYOUT. New Jersey, Pearson Prentice Hall, 2006. 648 p.
3. Huntington H. B., Grone A. R. Current-induced marker motion in gold wires. Journal of Physics and Chemistry of Solids, 1961, vol. 20, pp. 76-87.
4. Verbruggen A. H. Fundamental questions in the theory of electromigration. IBM Journal of. Research and Development, 1988, no. 32, pp. 93-98.
5. Tan C. M., Roy Arijit. Electromigration in ULSI Interconnection. Materials Science and Engineering. Elsevier, 2007, iss. 58, pp. 1-75.
6. Tao J., Young K. K., Cheung N. W., Hu C. Comparison of Electromigration Reliability of Tungsten and Aluminum Vias Under DC and Time-Varying Current Stressing. Proc. International Reliability Physics Symposium, 1992, vol. 14, no. 4, pp. 338-343.
7. Ames I., d'Heurle F. M., Horstmann R. E. Reduction of Electromigration in Aluminium Film by Copper Doping. IBM Journal of. Research and Development, 1970, vol.14, no. 4, pp. 461-463.
8. Iyer S. S., Ting Ch. Y. Electromigration Study of Al-Cu/Ti/Al-Cu System. Proc. International Reliability Physics Symposium, 1984, pp. 273-278.
9. Lloyd J. R., Smith P. M. The Effect of Passivation on the Electromigration Lifetime of Al/Cu Thin Film Conductors. Journal of Vacuum Science Technology A, 1983, vol. 1, no. 2, pp. 455-458.
10. Chen J. Z., Amerasekera A., Duvvury Ch. Design Methodology for Optimizing Gate Driven ESD Protection Circuits in Submicron CMOS Processes. *Proc. EOS/ESD Symposium*, 1997, pp. 1-10.
11. Lienig J, Thiele M. *Fundamentals of Electromigration-Aware Integrated Circuit Design*.1st Ed. SpringerLink, 2018, 159 p.
12. Weste, N. H. W., Harris D. M. CMOS VLSI Design: *A Circuits and Systems Perspective*. 4nd Ed. Boston. Addison-Wesley, Pearson Education, Inc., 2011, 838 p.
13. Gabriel C. T. Gate Oxide Damage: A Brief History and a Look Ahead. *Proceedings 6th International Symposium on Plasma Process-Induced Damage*, 2001, pp. 20-24.
14. Watanabe T., Yoshida Y. Dielectric Break-down of Gate Insulator Due to Reactive Ion Etching *Solid State Technology*, 1984, vol. 27, no. 44, pp. 263-266.
15. Mocuta A. C., Hook T. B., Chou A. I., Wagner T., Stamper A. K., Khare M., Gambino J. P. Plasma Charging Damage in SOI Technology. *Proceedings 6th International Symposium on Plasma Process-Induced Damage*, 2001, pp. 104-107.
16. Simon P., Luchies J.-M., Maly W. Antenna Ratio Definition for VLSI Circuits. *Proc. 4th International Symposium on Plasma Process-Induced Damage*, 1999, pp. 16-20.
17. Baker R. Jacob. *CMOS: Circuit Design, Layout, and Simulation*. 3nd Ed. New Jersey, John Wiley & Sons Inc., 2011, 1208 p.

Vladimir V. Shubin

Cand. of Sci. (Engineering), assistant professor of the Department of Technical Electronics of SibSUTIS, Head of the Development of analog ICs of JSC "NZPP Vostok department" (630082, Novosibirsk, Dachnaya str., 60), e-mail: shubin@nzpp.ru, ORCID ID: 0000-0002-2974-0497.

DOI: 10.55648/1998-6920-2024-18-3-73-85 УДК 658.562.012.7

Модификация метода идентификации и прослеживаемости сложных технических изделий с применением алгоритмов комбинаторной генерации на основе деревьев И/ИЛИ *

А. В. Токарева, Д. В. Кручинин

Томский государственный университет систем управления и радиоэлектроники (ТУСУР)

Аннотация: С развитием информационных технологий автоматизация производственных, технологических, управленческих процессов становится неотъемлемой частью функционирования предприятий. Это повышает не только качество производимых изделий, но и способствует оперативному устранению проблем на управленческом уровне, «прозрачности» в деятельности предприятия. Одними из систем, решающих задачу оперативного управления качеством изделий, являются системы прослеживаемости и идентификации. В статье отражена модификация оригинального метода идентификации и прослеживаемости с использованием алгоритмов комбинаторной генерации на основе деревьев И/ИЛИ. Метод предполагает представление информации о типе изделия в виде дерева И/ИЛИ по подхарактеристикам с дальнейшей возможностью генерации идентификационного номера типа изделия. Полученный метод обладает научной новизной, которая заключается в гибкости системы за счет выделения поддеревьев с информацией о комплектующих для заполнения предприятиям-поставщикам.

Ключевые слова: комбинаторная генерация, прослеживаемость, идентификация, деревья И/ИЛИ, менеджмент качества, сложное техническое изделие, автоматизация, ранжирование.

Для цитирования: Токарева А. В., Кручинин Д. В. Модификация метода идентификации и прослеживаемости сложных технических изделий с применением алгоритмов комбинаторной генерации на основе деревьев И/ИЛИ // Вестник СибГУТИ. 2024. Т. 18, № 3. С. 73–85. https://doi.org/10.55648/1998-6920-2024-18-3-73-85.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Токарева А. В., Кручинин Д. В., 2024

Статья поступила в редакцию 23.01.2024;
переработанный вариант – 21.03.2024;
принята к публикации 31.03.2024.

1. Введение

В настоящее время автоматизация производственных, технологических и управленческих процессов играет важную роль в деятельности предприятия. Прежде всего это связано с разнообразием технических решений в виде программного обеспечения, а также необходимостью своевременного контроля всех аспектов работы предприятия.

Одним из направлений, в котором широко применяются новые технологии, является менеджмент качества продукции. Понятие «менеджмент качества» согласно ГОСТ ИСО 9000-2015 – это менеджмент применительно к качеству [1]. Данное понятие является комплекс-

*Работа выполнена в рамках гранта Российского научного фонда (проект № 22-71-10052).

ным и включает в себя разработку политики, целей в области качества посредством планирования, обеспечения, управления и улучшения процессов на предприятии. Одним из элементов менеджмента качества является качество производимой продукции. Создание продукта – это многоэтапный процесс, который связан с большим количеством факторов. Поэтому качество выпускаемой продукции является ключевым показателем работы предприятия, а современное понятие «качество» включает в себя не конечное мнение ключевого потребителя о продукте, а более сложную структуру, которая подразумевает под собой качество управления, конечного продукта, поставок и работы сотрудников, а также унификацию производственных процессов.

Предприятия часто сталкиваются с многочисленными проблемами, связанными с недостатком информации о том, что происходит с продукцией во время сборки и производства. Из-за отсутствия актуальной информации о том, на каком этапе производства находится партия изделий, какие материалы и комплектующие были использованы, могут возникнуть негативные последствия: нарушение технологии производства, сбои в планировании закупок, невозможность управления жизненным циклом продукта. Такая специфика проблем характерна для предприятий, производящих сложные технические изделия, отличающиеся продолжительным временем создания, использованием наукоемких технологий, большим количеством поставщиков комплектующих. Решением указанных проблем становится внедрение системы принципов идентификации и прослеживаемости продукции, которая в настоящее время является обязательным и основным элементом системы менеджмента качества применительно к продукции.

Принцип идентификации подразумевает присвоение каждому продукту уникального признака, по которому его можно отличить от остальных. Особенностями данного принципа является нацеленность на конечный результат благодаря установлению тождественности, систематизации и кластеризации производимой на предприятии продукции.

Принцип прослеживаемости подразумевает возможность отслеживания движения, местонахождения и происхождения продукции на всех стадиях производства, обработки и распределения. Движение продукции может относиться к происхождению сырья, истории обработки или ее распределения [2].

Согласно исследованиям современная концепция идентификации и прослеживаемости продукции появилась вследствие начала применения высоких технологий в различных сферах деятельности [3]. Изначально особое внимание уделялось таким сферам производства, как пищевая промышленность, сельское хозяйство, тяжелая промышленность. В настоящее время данные принципы автоматизированы, и независимо от производства на предприятиях применяются автоматизированные системы идентификации и прослеживаемости (СИиП).

Говоря об определении системы прослеживаемости и идентификации как программного средства, В. А. Носенко [4] отмечает, что это автоматизированная система, которая позволяет собирать, контролировать и анализировать информацию о технологических операциях, персонале и ресурсах, задействованных в процессе производства. Основным отличием данного вида программного обеспечения является сложность структуры в связи с необходимостью обеспечения строгой взаимосвязи между информационным (данные о сырье, комплектующих и т.д.) и физическим (непосредственно комплектующие, сырье и т.д.) потоками.

В большей части таких систем применяется технология «блокчейн», которая описана рядом зарубежных и отечественных авторов [5–8]. Данная технология подразумевает представление производства в виде «цепочки» последовательных процессов, которые позволяют отследить все этапы создания изделия, а также отобразить данные, напрямую связанные с самим изделием.

С точки зрения идентификации в настоящее время практически все производства используют несколько основных способов идентификации: идентификационный номер, штриховое линейное и двумерное кодирование, радиочастотные метки. Набор инструментов для генерации идентификаторов обширен: это могут быть интегрированные в систему программы,

которые автоматически создают штрих-коды, отдельные библиотеки в языках программирования и даже генераторы открытого доступа в сети Интернет [9–13].

В связи с тем, что внедрение принципов прослеживаемости и идентификации на предприятии является обязательным условием для их деятельности, автоматизированная система не только повышает эффективность применения данных принципов и делает процесс производства «прозрачным», но также позволяет оперативно принимать решения по возникающим внештатным ситуациям.

2. Цель и методы исследования

В настоящее время рынок отечественных систем идентификации и прослеживаемости представлен большим количеством программ. Из наиболее крупных можно выделить программные продукты компаний Клеверенс, ДатаПермь, 1С, Честный знак. При этом основной набор выполняемых функций обширен и практически аналогичен:

– идентификация поступающих на предприятие материалов и комплектующих (обеспечение привязки поставок материалов и комплектующих к системе хранения и перемещения предприятия);

– прослеживание использования промаркированных материалов и комплектующих в процессе изготовления продукции;

– обеспечение информацией для своевременного изъятия из производства изделий и материалов, которые не удовлетворяют установленным требованиям;

– идентификация отдельного изделия (выбор технологии маркировки, генерация штрихкодов, подготовка для нанесения, возможность отображения информации об уникальном изделии);

– отслеживание использования готовой продукции (возможность контроля пути изделия от создания до реализации конечному пользователю).

Однако у таких систем есть ряд проблем: отсутствие универсальности (система настраивается под конкретное производство), большой объем хранимой и динамически растущей информации, не все виды идентификации позволяют хранить полный объем информации об изделии как для работников, так и для покупателя. Решение первой проблемы находится в настройке программного обеспечения под нужды предприятия, однако вопросы, связанные с хранением информации об изделии и отображением через идентификатор необходимой информации, возможно решить через внедрение нового математического аппарата.

Целью данного исследования является модификация метода идентификации и прослеживаемости сложных технических изделий с использованием алгоритмов комбинаторной генерации, основанных на деревьях И/ИЛИ.

Как самостоятельное направление комбинаторная генерация описана Ф. Раски [14], где выделены следующие четыре задачи над комбинаторными множествами:

– listing – последовательная генерация элементов рассматриваемого комбинаторного множества;

– ranking – ранжирование (нумерация) элементов рассматриваемого комбинаторного множества;

– unranking – генерация элементов рассматриваемого комбинаторного множества в соответствии с их рангами (номерами);

– random selection – генерация элементов рассматриваемого комбинаторного множества в случайном порядке.

Понятие дерева И/ИЛИ было введено Д. Слейглом [15] для построения системы взятия интегралов в символьном виде. Подробно сущность деревьев И/ИЛИ описана в работах В. В. Кручинина [16,17] и подразумевает под собой древовидную структуру представления множества, которая содержит узлы двух типов: И-узел и ИЛИ-узел. Графическое представление И-узла отличается от ИЛИ-узла тем, что все ветви, ведущие к потомкам И-узла, объ-

единяются дугой. В свою очередь, вариантом дерева И/ИЛИ называется дерево, которое получается из данного путем отсечения всех дуг, кроме одной, у всех ИЛИ-узлов (рис. 1).

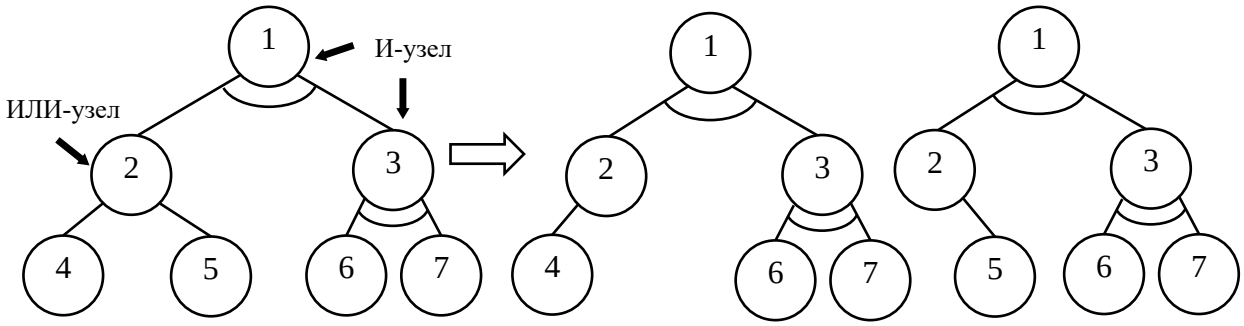


Рис. 1. Пример дерева И/ИЛИ и всех его вариантов

При этом существуют ограничения, влияющие на возможность применения алгоритмов комбинаторной генерации с использованием структур деревьев И/ИЛИ: функция мощности рассматриваемого комбинаторного множества должна принадлежать алгебре $\{\mathbb{N}, +, \times, R\}$, где R – оператор примитивной рекурсии. В основе логики работы алгоритмов ранжирования и генерации по рангу вариантов дерева И/ИЛИ лежит функция $l(z)$, которая вычисляет значение ранга поддерева для заданного узла z. Правила вычисления значения $l(z)$ зависят от типа узла z:

– если z является листом, то $l(z) = 0$;

– если z является ИЛИ-узлом, то $l(z) = l\left(s_k^{(z)}\right) + \sum_{i=1}^{k-1} w\left(s_i^{(z)}\right)$, где $w\left(s_i^{(z)}\right)$ показывает количество всех вариантов для поддерева i-го потомка узла z, а k показывает порядковый номер выбранного в варианте потомка узла z среди всех n его потомков;

– если z является И-узлом, то $l(z) = l\left(s_1^{(z)}\right) + w\left(s_1^{(z)}\right) \cdot \left(\ldots\left(l\left(s_{n-1}^{(z)}\right) + w\left(s_{n-1}^{(z)}\right) l\left(s_n^{(z)}\right)\right)\ldots\right)$.

Тогда алгоритм ранжирования варианта дерева И/ИЛИ представляет собой вычисление значения $l(z)$, где z является корневым узлом структуры дерева И/ИЛИ. Алгоритм генерации варианта дерева И/ИЛИ по его рангу базируется на обратных преобразованиях.

Метод идентификации и прослеживаемости, описанный в работе [17], позволяет представить всю информацию об изделии в виде дерева И/ИЛИ и с помощью алгоритма Rank сгенерировать идентификационный номер, являющийся ссылкой на данные об изделии (рис. 2). Под изделием в данном методе подразумевается тип изделия, то есть совокупность производимых изделий, обладающих идентичными характеристиками. В случае необходимости получения информации по конкретному изделию можно восстановить ветку дерева с полным описанием, используя алгоритм Unrank. Прослеживаемость в данном методе представляет собой возможность представить информацию об изделии в виде цепочки взаимосвязанной информации, начиная с использованных комплектующих и заканчивая серийным номером конкретного готового изделия.

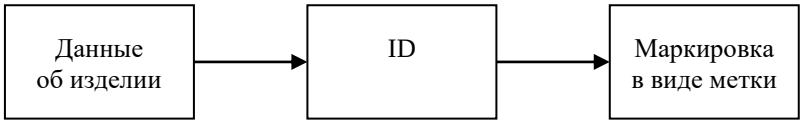


Рис. 2. Этапы генерации идентификатора

Однако данный метод подходит для фиксированных множеств, а существующая программная реализация метода устарела. Кроме того, нанесение полученного идентификаци-

онного номера не всегда актуально, особенно на серийных производствах с большими объемами производимых изделий, так как в некоторых сферах в обязательном порядке используется маркировка с применением штрихового кодирования или радиочастотных меток.

3. Основные результаты

Предлагается следующая модификация оригинального метода, отличительными особенностями которого являются гибкость системы идентификации и прослеживаемости за счет предоставления возможности на заполнение информации о комплектующих изделия предприятиям-поставщикам и определение резервных позиций в справочниках характеристик.

Шаг 1. Определение характеристик рассматриваемых изделий.

Для того, чтобы представить изделия в виде структуры дерева И/ИЛИ, необходимо определить множество характеристик изделий (под характеристиками понимаются комплектующие изделия и параметры, идентифицирующие полное представление изделия). Помимо этого, определяются взаимосвязи между характеристиками изделий (в изделии используется либо одна из перечня возможных характеристик, либо их комбинация).

Шаг 2. Представление модели изделий в виде структуры дерева И/ИЛИ.

Каждая характеристика изделий представляется ИЛИ-узлом, если подразумевается использование одной из перечня возможных ее подхарактеристик, либо И-узлом, если подразумевается использование комбинации всех ее подхарактеристик. Лист дерева – это характеристика, которая далее не декомпозируется. Также при формировании дерева И/ИЛИ создаются справочники для каждого из узлов, которые позволяют раскрыть информацию об их потомках, то есть содержат подробную информацию о каждой характеристике изделий.

Шаг 3. Определение резервных позиций в справочниках характеристик.

Под резервными позициями понимаются потомки ИЛИ-узлов, которые изначально не содержат описание какой-либо подхарактеристики и, следовательно, обладают пустой записью в соответствующем справочнике характеристик (рис. 3).

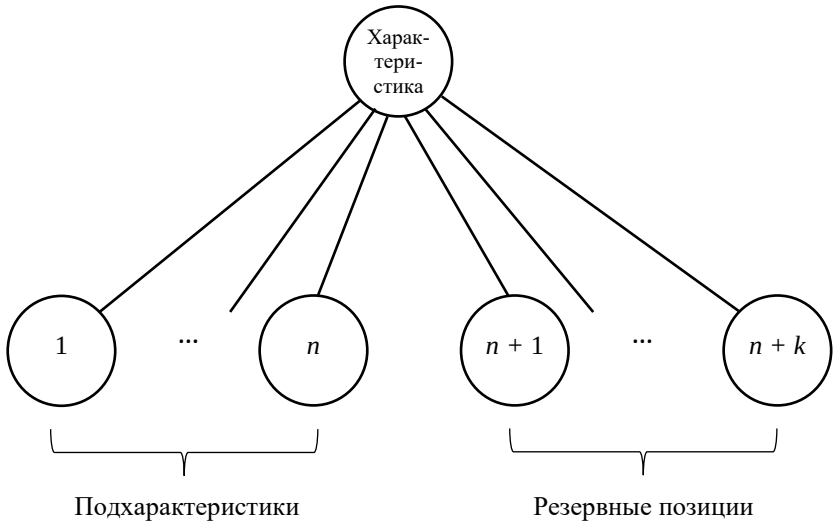


Рис. 3. Резервные позиции в справочниках характеристик

В дальнейшем такие резервные позиции в справочниках характеристик можно заполнить необходимой новой информацией, при этом структура дерева И/ИЛИ и количество его вариантов не изменится, пример справочника приведен в табл. 1. Количество резервных позиций определяется индивидуально по потребностям предприятия, например, в зависимости от частоты обновления поставщиков.

Таблица 1. Представление справочника

1	Описание подхарактеристики 1
2	Описание подхарактеристики 2
…	…
n	Описание подхарактеристики *n*
n + 1	Резерв 1
…	…
n + *k*	Резерв *k*

Шаг 4. Определение характеристик, предоставляемых предприятиями-поставщиками.

На данном шаге определяются необходимые для прослеживаемости и идентификации изделия характеристики от предприятий-поставщиков комплектующих. Таким образом, каждое предприятие-поставщик выполняет Шаги 1–3 данного метода в рамках поддерева узла предоставленной ему характеристики. В свою очередь, это позволяет предприятию не углубляться самостоятельно в декомпозицию отдельной характеристики, делегировав выполнение задачи специализирующемуся на данной характеристике предприятию-поставщику. В зависимости от количества разных предприятий-поставщиков, предоставляющих информацию о поддереве узла одной характеристики, возможны следующие ситуации:

1) для рассматриваемой характеристики имеется дерево $D(P)$ от единственного предприятия-поставщика, тогда корень дерева $D(P)$ становится потомком узла, соответствующего данной характеристике (рис. 4);

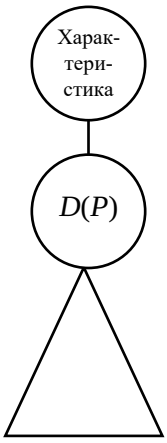


Рис. 4. Представление информации о характеристике от одного предприятия-поставщика

2) для рассматриваемой характеристики имеется *m* деревьев $D(P_1), D(P_2), ..., D(P_m)$ от *m* предприятий-поставщиков, тогда узел, соответствующий данной характеристике, представляется И-узлом с двумя потомками, первый из которых – ИЛИ-узел с информацией о выбранном предприятии-поставщике P_i, а второй – поддерево $D(P_i)$ с информацией о характеристике от выбранного предприятия-поставщика P_i (рис. 5).

В данном случае деревья $D(P_1), D(P_2), ..., D(P_m)$ предприятий-поставщиков с информацией о характеристике могут иметь разный вид. Однако для корректной работы алгоритмов комбинаторной генерации на основе деревьев И/ИЛИ деревья $D(P_i)$ должны иметь одинаковое количество вариантов, то есть $w(D(P_1)) = w(D(P_2)) = ... = w(D(P_m))$.

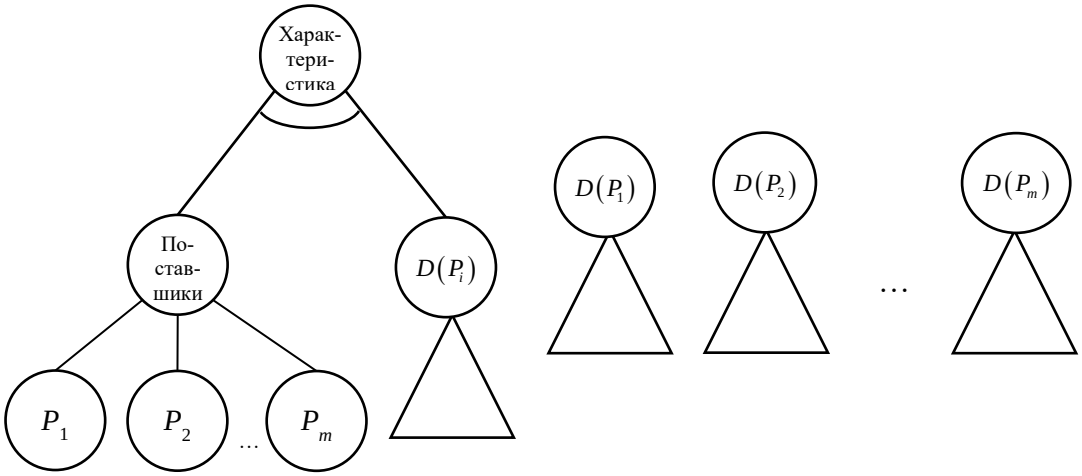


Рис. 5. Представление информации о характеристике от *m* предприятий-поставщиков

Если данное условие не выполняется, то необходимо выполнить следующее преобразование структуры дерева И/ИЛИ:

– находим дерево с максимальным количеством вариантов $M = \max_i \left(w\left(D\left(P_i \right) \right) \right)$;

– для каждого дерева $D\left(P_i \right)$: если $w\left(D\left(P_i \right) \right) < M$, то взамен дерева $D\left(P_i \right)$ строится дерево $D^*\left(P_i \right)$, содержащее *M* вариантов за счет добавления $M - w\left(D\left(P_i \right) \right)$ фиктивных потомков-листов (рис. 6).

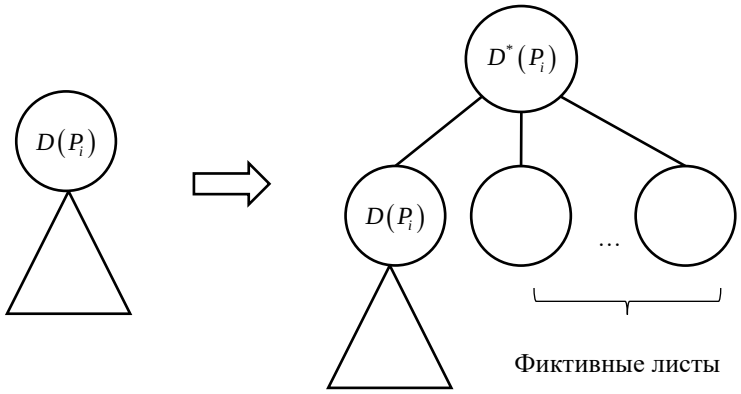


Рис. 6. Добавление фиктивных узлов

Шаг 5. Сбор общей структуры дерева И/ИЛИ.

Композиция дерева И/ИЛИ от предприятия и деревьев И/ИЛИ от предприятий-поставщиков с добавлением всех требуемых резервных и фиктивных позиций формируют общую структуру дерева И/ИЛИ для модели рассматриваемых изделий. Общая структура дерева И/ИЛИ представляет собой совокупность всех теоретически возможных вариантов производства изделий, при этом конкретный тип изделия можно получить путем построения одного варианта дерева И/ИЛИ.

Шаг 6. Применение алгоритма Rank для генерации идентификационного номера.

Сопоставив заданное изделие с одним из вариантов дерева И/ИЛИ, можно сгенерировать идентификационный номер изделия с помощью алгоритма комбинаторной генерации Rank. Сгенерированный идентификатор заносится в базу данных и наносится на изделие в виде некоторой метки, которая может быть представлена как в виде нанесенного номера, так и в формате двумерного штрихового кода (QR-код).

Если возникает ситуация, что приборы имеют одинаковый идентификатор (то есть идентичны по всем характеристикам и относятся к одному и тому же типу изделия), можно использовать сгенерированный ранг в сочетании с порядковым номером прибора.

Шаг 7. Применение алгоритма Unrank для получения информации об изделии.

При необходимости получения информации по конкретному изделию с помощью алгоритма комбинаторной генерации Unrank по идентификационному номеру генерируется вариант дерева И/ИЛИ со всеми характеристиками данного изделия.

Шаг 8. Интерпретация полученных результатов.

Выполняется представление полученной информации по конкретному изделию в понятном пользователю виде путем сопоставления варианта дерева И/ИЛИ с данными из справочников характеристик изделий. Так как из-за большого количества характеристик вариант дерева может быть большим и неудобным для чтения, то информацию о нем можно представить в виде таблицы.

Преимущества предложенного модифицированного метода по сравнению с оригинальным заключаются в следующем:

1. Дает гибкость системы идентификации и прослеживаемости с помощью разграничения узлов основного предприятия и предприятий-поставщиков комплектующих.

2. Позволяет хранить и собирать больше информации об изделии без изменения структуры дерева И/ИЛИ благодаря заложенному в нем резерву, что влияет на точность и оперативность принятия решений в случае нештатных ситуаций.

3. Уменьшает объем хранимой информации за счет использования справочников-классификаторов.

4. Позволяет разграничивать доступ к информации лицам, принимающим решения.

Применение данного метода в качестве основы программного модуля для автоматизированных систем управления подходит для производств, на которых создаются сложные технические изделия, обладающие большим количеством комплектующих и продолжительным процессом создания (например, ракетостроение, электроника и другие наукоемкие предприятия). При этом с точки зрения программной реализации для хранения и обработки структуры дерева И/ИЛИ можно использовать разные структуры данных. Одним из таких вариантов является реализация самостоятельных структур данных для дерева И/ИЛИ и для варианта дерева И/ИЛИ. Тогда в оперативной памяти структура дерева И/ИЛИ и его варианты будут храниться как экземпляры соответствующего класса, а их хранение в постоянной памяти организовывается путем сериализации экземпляров классов. В качестве еще одного способа представления структуры дерева И/ИЛИ является организация структуры связного списка списков, где каждый узел дерева И/ИЛИ содержит информацию о типе узла и список указателей на потомков данного узла. Тогда в постоянной памяти структура дерева И/ИЛИ и его варианты могут храниться как текстовая строка правильной скобочной последовательности, где каждая пара открытой и закрытой скобок соответствует одному узлу дерева И/ИЛИ, а вложенность скобок описывает иерархическую структуру узлов дерева И/ИЛИ. Кроме того, структуре дерева И/ИЛИ можно однозначно сопоставить структуру реляционной базы данных, а совокупность связанных записей такой базы данных будет соответствовать конкретному варианту дерева И/ИЛИ. Использование последнего способа представления структуры дерева И/ИЛИ и его вариантов позволяет эффективно организовать процессы их обработки, в том числе поиск вариантов дерева И/ИЛИ с заданными характеристиками, за счет наличия соответствующего функционала в области работы с реляционными базами данных.

Представим применение предложенного метода на абстрактном сложном техническом изделии. Допустим, на предприятии производятся приборы, которые имеют n деталей в верхнем И-узле. Рассмотрим дочерний И-узел «Деталь 1». «Деталь 1» включает в себя две основных детали и технологию производства. Потомок И-узла «Деталь 1.1» представляет собой И-узел с информацией об этой детали от предприятия-поставщика. При этом левый потомок (ИЛИ-узел «Поставщик») содержит в себе конечные листья с наименованиями поставщиков детали, а правый узел (узел «1») передается поставщику на самостоятельное заполнение. Модель полученной структуры дерева И/ИЛИ представлена на рис. 7.

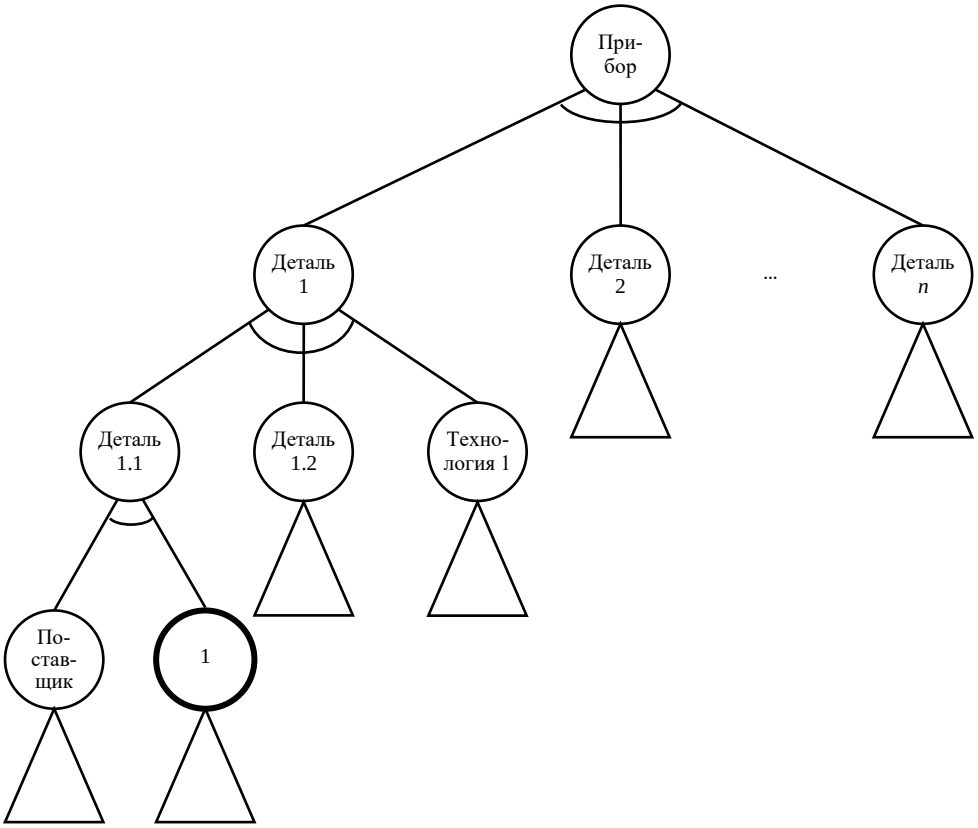


Рис. 7. Представление информации о приборе в виде дерева И/ИЛИ

Узел «1», выделяемый для заполнения предприятием-поставщиком, может иметь разнообразную структуру. Примеры таких структур представлены на рис. 8.

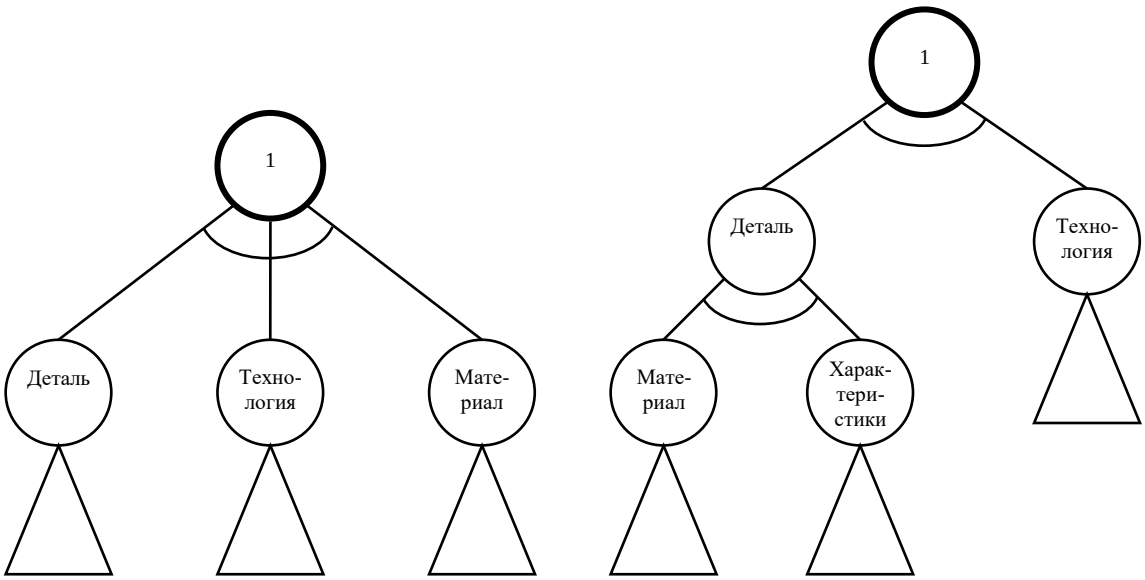


Рис. 8. Возможные варианты структуры узла от предприятия-поставщика

При заданном дереве И/ИЛИ с информацией о приборе можно выполнить алгоритм Rank и получить идентификационный номер прибора, который возможно использовать как в качестве нанесенного на прибор идентификатора, так и для занесения в базу данных. При имеющемся идентификаторе выполняется обратный алгоритм Unrank и строится вариант дерева с данными об этом приборе.

4. Заключение

Управление качеством производимой продукции во многом зависит от комплектующих, стандартизации технологических процессов, возможности отследить путь изделия от производства до реализации. В настоящее время на предприятиях автоматизированы и внедрены в обязательном порядке принципы идентификации и прослеживаемости. Это позволяет повысить оперативность принятия решения в случае нештатных ситуаций, а также сделать процесс производства более прозрачным.

В данной работе предложен модифицированный метод идентификации и прослеживаемости изделий, основанный на использовании алгоритмов комбинаторной генерации на основе деревьев И/ИЛИ, а также представлен пример модели древовидной структуры для абстрактного сложного технического изделия. Преимущества использования данного метода заключаются в гибкости системы идентификации и прослеживаемости, уменьшении объема хранимой информации. Применение данного метода возможно в дальнейшем в качестве одного из программных модулей автоматизированных систем управления, а точнее систем идентификации и прослеживаемости продукции.

Литература

1. ГОСТ Р ИСО 9000-2015: Национальный стандарт Российской Федерации «Системы менеджмента качества». Москва, 2015.
2. *Савинкова Е. А., Гусева И. В., Пиркина О. В.* Идентификация и прослеживаемость продукции как одно из основных требований при внедрении системы управления качеством на примере предприятия ООО «Шик» Чувашской республики // Вестник Марийского государственного университета. Серия «Сельскохозяйственные науки. Экономические науки». 2016. Т. 2, № 4. С. 40–43.
3. *Дравица В., Курбацкий А.* Промышленная революция Industry 4.0 // Наука и инновации. 2016. № 3. С. 13–16.
4. *Носенко В. А., Силаев А. А., Евремкин С. И., Гредников С. Б.* Исследование инструментов идентификации автоматизированной системы прослеживаемости сборочного производства // Сборник материалов X Международной научно-технической конференции «Инновации в машиностроении». Кемерово, 26–29 ноября, 2019. С. 199–204.
5. *Wang S., Dongyl L., Zhang Y., Chen J.* Smart contract-based product traceability system in the supply chain scenario // IEEE Access. 2019. V. 7. P. 115122–115133.
6. *Турсунходжаев М. Л., Загребельская М. В.* Некоторые аспекты использования блокчейн-технологии в целях оптимизации цепочек поставок // Трансформация моделей корпоративного управления в условиях цифровой экономики. 2022. № 1. С. 1–4.
7. *Hong W., Cai Y., Yu Z., Yu X.* An agri-product traceability system based on IoT and blockchain technology // Proc. IEEE International Conference on Hot Information-Centric Networking (HotICN), Shenzhen, China, August 15–17, 2018. P. 254–255.
8. *Тимчук Е. Г.* Применение технологии блокчейн в целях обеспечения прослеживаемости пищевой продукции: текущее состояние и перспективы // Научные труды Дальрыбвтуза. 2022. Т. 61, № 3. С. 13–20.
9. *Коробко И.* Генерация штрих-кода EAN-13 // Системный администратор. 2012. № 5. С. 74–78.
10. *Стариков А. В., Стародубцева Т. Н., Батурин К. В., Поляков С. Ю.* Автоматизация идентификации и учета деталей мебельного производства с использованием системы штрихового кодирования // Лесотехнический журнал. 2013. № 2. С. 138–149.
11. *Карев В. А., Кравец А. Г.* Разработка алгоритма генерации QR-кода для онлайн платежей // Системный анализ в науке и образовании. 2019. № 1. С. 1–6.

12. *Абрамович В. В., Брумштейн Ю. М., Шипилова О. В.* Двумерные графические коды: анализ информационной емкости, способов генерации, направлений применения // Прикаспийский журнал: управление и высокие технологии. 2019. № 1. С. 10–33.
13. *Абрамович В. В., Брумштейн Ю. М.* Одномерные (линейные) графические коды: анализ способов генерации, традиционных и новых направлений применения, вопросов информационной безопасности использования // Прикаспийский журнал: управление и высокие технологии. 2018. № 3. С. 19–36.
14. *Ruskey F.* Combinatorial generation. Working version (1j-CSC 425/520). 2003. 311 p. URL: https://page.math.tu-berlin.de/~felsner/SemWS17-18/Ruskey-Comb-Gen.pdf (дата обращения: 20.01.2024).
15. *Slagle J.* A heuristic program that solves symbolic integration problem in freshman calculus // Journal of the ACM. 1963. V. 10, № 4. P. 507–520.
16. *Кручинин В. В., Люкшин Б. А.* Метод кодирования информационных объектов на основе деревьев И/ИЛИ // Доклады ТУСУР. 2010. № 1. С. 170–172.
17. *Кручинин В. В*. Методы, алгоритмы и программное обеспечение комбинаторной генерации: дис. ... д-ра техн. наук: 05.13.11. Томск, 2010. 387 с.

Токарева Алина Вячеславовна
младший научный сотрудник лаборатории алгоритмов и технологий исследования дискретных структур, Томский государственный университет систем управления и радиоэлектроники (ТУСУР, 634050, Томск, пр. Ленина, д. 40), e-mail: alina.v.tokareva@tusur.ru, ORCID ID: 0000-0003-1353-0177.

Кручинин Дмитрий Владимирович
д.т.н., профессор кафедры компьютерных систем в управлении и проектировании, Томский государственный университет систем управления и радиоэлектроники (ТУСУР, 634050, Томск, пр. Ленина, д. 40), e-mail: kdv@fb.tusur.ru, ORCID ID: 0000-0003-3412-432X.

Авторы прочитали и одобрили окончательный вариант рукописи.
Авторы заявляют об отсутствии конфликта интересов.
Вклад соавторов: *Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.*

Modification of the Method for Identifying and Tracing Complex Technical Products Using Combinatorial Generation Algorithms Based on AND/OR Trees

Alina V. Tokareva, Dmitry V. Kruchinin

Tomsk State University of Control Systems and Radioelectronics (TUSUR)

Abstract: The development of information technologies, automation of production, technological, and management processes are becoming part of production enterprises. This increases not only the quality of replacement products, but also the ability to resolve problems promptly at the management level and «transparency» in the activities of the enterprise. One of the systems that defines the task of operational product quality management is traceability and identification systems. The article edits a modification of the original identification and traceability method using combinatorial generation algorithms based on AND/OR calculations. The method provides information about the type of product in the form of an and/or tree by subcharacteristics, followed by obtaining a definition of the product type number. A method obtained has scientific novelty consisting of a flexible system for identifying subtrees with information about components to be filled in by supplier enterprises.

Keywords: combinatorial generation, traceability, identification, AND/OR trees, quality management, complex technical product, automation, ranking.

For citation: Tokareva A. V., Kruchinin D. V. Modification of the method for identifying and tracing complex technical products using combinatorial generation algorithms based on AND/OR trees (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 3, pp. 73-85. https://doi.org/10.55648/1998-6920-2024-18-3-73-85.

Content is available under the license Creative Commons Attribution 4.0 License

© Tokareva A. V., Kruchinin D. V., 2024

The article was submitted: 23.01.2024;
revised version: 21.03.2024;
accepted for publication 31.03.2024.

References

1. *GOST R ISO 9000-2015. Sistemy menedzhmenta kachestva. Osnovnye polozhenija i slovar*' [Russian Standard No. 9000-2015 Quality management systems. Fundamentals and vocabulary]. Moscow, 2015, p. 53.
2. Savinkova E. A., Guseva I. V., Pirkina O. V. Identifikacija i proslezhivaemost' produkcii kak odno iz osnovnyh trebovanij pri vnedrenii sistemy upravlenija kachestvom na primere predprijatija OOO «Shik» Chuvashskoj respubliki [Identification and traceability of products as one of the main requirements for a quality management system implementation on the example of LLC “SHIK” in the Republic of Chuvashia]. *Vestnik Marijskogo gosudarstvennogo universiteta*, 2016, no. 4 (8), pp. 40-43.
3. Dravica V., Kurbackij A. Promyshlennaja revoljucija Industry 4.0 [Industrial revolution Industry 4.0]. *Nauka i innovacii*, 2016, no. 3, pp. 13-16.
4. Nosenko V. A., Silaev A. A., Evremkin S. I., Grednikov S. B. Issledovanie instrumentov identifikacii avtomatizirovannoj sistemy proslezhivaemosti sborochnogo proizvodstva [Research of tools of identification used at automated traceability system for assembly manufacturing]. *Sbornik materialov X Mezhdunarodnoj nauchno-tehnicheskoj konferencii «Innovacii v mashinostroenii*, 2019, pp. 199-204.
5. Wang S., Dongyl L., Zhang Y., Chen J. Smart contract-based product traceability system in the supply chain scenario. *IEEE Access*, 2019, vol. 7. pp. 115122–115133.
6. Tursunhodzhaev M. L., Zagrebel'skaja M. V. Nekotorye aspekty ispol'zovanija blokchejn-tehnologii v celjah optimizacii cepochek postavok [Some aspects of using blockchain technology to optimize supply chains]. *Transformacija modelej korporativnogo upravlenija v uslovijah cifrovoj ekonomiki*, 2022, no. 1, pp. 1-4.
7. Hong W., Cai Y., Yu Z., Yu X. An agri-product traceability system based on IoT and blockchain technology. *Proc. of 2018 IEEE International Conference on Hot Information-Centric Networking (HotICN)*, 2019, pp. 254-255.
8. Timchuk E. G. Primenenie tehnologii blokchejn v celjah obespechenija proslezhivaemosti pishhevoj produkcii: tekushhee sostojanie i perspektivy [Application of blockchain technology in order to ensure the traceability of food products: current state and prospects]. *Nauchnye trudy Dal'rybvtuza*, 2022, vol. 61, no 3, pp. 13-20.
9. Korobko I. Generacija shtrih-koda EAN-13 [EAN-13 barcode generation]. *System Administrator*, 2012, no. 5, pp. 74-78.
10. Starikov A. V., Starodubceva T. N., Baturin K. V., Poljakov S. Ju. Avtomatizacija identifikacii i ucheta detalej mebel'nogo proizvodstva s ispol'zovaniem sistemy shtrihovogo kodirovanija [Automation of identification and accounting of furniture production parts using a bar coding system]. *Lesotehnicheskij zhurnal*, 2013, no. 2, pp. 138-149.
11. Karev V. A., Kravec A. G. Razrabotka algoritma generacii QR-koda dlja onlajn platezhej [Development of QR-code generation algorithm for online payments]. *Setevoe nauchnoe izdanie «Sistemnyj analiz v nauke i obrazovanii»*, 2019, no. 1, pp. 1-6.
12. Abramovich V. V., Brumshtejn Ju. M., Shipilova O. V. Dvumernye graficheskie kody: analiz informacionnoj emkosti, sposobov generacii, napravlenij primenenija [Two-dimensional graphic codes: anal-

ysis of information capacity, ways of generation, directions of application]. *Prikaspijskij zhurnal: upravlenie i vysokie tehnologii*, 2019, no. 1, pp. 10-33.
13. Abramovich V. V., Brumshtejn Ju. M. Odnomernye (linejnye) graficheskie kody: analiz sposobov generacii, tradicionnyh i novyh napravlenij primenenija, voprosov informacionnoj bezopasnosti ispol'zovanija [One-dimensional (linear) graphic codes: analysis of ways of generation, traditional and new directions of application, questions of information security of use]. *Prikaspijskij zhurnal: upravlenie i vysokie tehnologii*, 2018, no. 3, pp. 19-36.
14. Ruskey F. *Combinatorial generation. Working version (1j-CSC 425/520)*, 2003, available at: https://page.math.tu-berlin.de/~felsner/SemWS17-18/Ruskey-Comb-Gen.pdf (accessed: 20.01.2024).
15. Slagle J. A heuristic program that solves symbolic integration problem in freshman calculus. *Journal of the ACM*, 1963. vol. 10, no. 4, pp. 507-520.
16. Kruchinin V. V., Ljukshin B. A. Metod kodirovanija informacionnyh ob`ektov na osnove derev'ev I/ILI [Method of Coding of Information Objects on the Basis of Trees And-Or]. *Doklady TUSUR*, 2010, no. 1, pp. 170-172.
17. Kruchinin V. V. *Metody, algoritmy i programmnoe obespechenie kombinatornoj generacii* [Methods, algorithms and software for combinatorial generation]. Doctor's degree dissertation, Tomsk, 2010. 387 p.

Alina V. Tokareva

Junior Researcher, Laboratory of Algorithms and Discrete Research Technologies, Tomsk State University of Control Systems and Radioelectronics (TUSUR, Russia, 634050, Tomsk, Lenin Ave., 40), e-mail: alina.v.tokareva@tusur.ru, ORCID ID: 0000-0003-1353-0177.

Dmitry V. Kruchinin

Dr. of Sci. (Engineering), Professor of the Department of Computer Systems in Management and Design, Tomsk State University of Control Systems and Radioelectronics (TUSUR, Russia, 634050, Tomsk, Lenin Ave., 40), e-mail: kdv@fb.tusur.ru, ORCID ID: 0000-0003-3412-432X.

DOI: 10.55648/1998-6920-2024-18-3-86-98 УДК 519.612

Сравнение программных реализаций методов решения СЛАУ в задаче нахождения равновесного состава сложной многокомпонентной гетерогенной системы

П. А. Сеченов, И. А. Рыбенко

Сибирский государственный индустриальный университет (СибГИУ), г. Новокузнецк

Аннотация: В статье представлен сравнительный анализ быстродействия программных реализаций численных методов при решении задачи нахождения равновесного состава сложной многокомпонентной гетерогенной системы. Задача нахождения равновесного состава системы подразделяется на следующие подзадачи: 1) учет ограничений (метод Лагранжа); 2) нахождение функции максимума нелинейной функции: 2.1) преобразование функции в систему линейных уравнений (метод Ньютона – Рафсона); 2.2) использование численных методов для решения системы линейных алгебраических уравнений (СЛАУ). Аналитический обзор литературных данных показал, что при решении систем линейных алгебраических уравнений градиентные методы имеют лучшее быстродействие. Поэтому в статье сравнивалось быстродействие программной реализации всего алгоритма при использовании прямых (Гаусса, LUP-разложение) и итерационных методов (метод сопряженных градиентов, метод стабилизированных бисопряженных градиентов) решения системы линейных алгебраических уравнений. Также было проведено сравнение скорости вычисления разработанной программы с использованием динамически подключаемых библиотек Alglib, ILNumerics, MathNet, Accord для решения СЛАУ в задаче нахождения равновесного состава термодинамической системы.

Ключевые слова: равновесный состав, программная реализация, численные методы, метод сопряженных градиентов, метод стабилизированных бисопряженных градиентов, быстродействие, динамически подключаемые библиотеки.

Для цитирования: Сеченов П. А., Рыбенко И. А. Сравнение программных реализаций методов решения СЛАУ в задаче нахождения равновесного состава сложной многокомпонентной гетерогенной системы // Вестник СибГУТИ. 2024. Т. 18, № 3. С. 86–98. https://doi.org/10.55648/1998-6920-2024-18-3-86-98.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Сеченов П. А., Рыбенко И. А., 2024

Статья поступила в редакцию 11.01.2024;
переработанный вариант – 25.02.2024;
принята к публикации 02.04.2024.

1. Введение

Одной из важных задач во многих областях науки является определение области допустимых значений термодинамических функций и параметров сложных многокомпонентных термодинамических систем, в которых изменения состояния сопровождаются фазовыми, полиморфными и химическими превращениями. Решить такую задачу можно только на основе полного термодинамического анализа, позволяющего определять равновесный состав системы по свойствам составляющих ее компонентов при заданных термодинамических парамет-

рах. Для этого предлагается использовать метод термодинамического моделирования, основанный на принципе максимума энтропии, который заключается в том, что в равновесном состоянии энтропия изолированной системы максимальна. Тогда задача определения равновесного состава сложной гетерогенной многокомпонентной системы сводится к определению максимума энтропии как функции состава:

$$S(U,V,n) \Rightarrow \max,$$

где S – энтропия, Дж / (моль ·К); U – полная внутренняя энергия, Дж / (моль ·К); V – объём, м^3; n – количество вещества, моль;

при следующих ограничениях:

- постоянство полной внутренней энергии системы;
- соблюдение условий материального баланса.

Задачу нахождения равновесного состава представим в виде решения системы нелинейных алгебраических уравнений с ограничениями, представленными выше. Подробное решение этой задачи рассматривалось ранее в статье [1] с использованием численных методов (рис. 1): 1) Лагранжа [2, 3] для перехода от задачи с ограничениями к задаче без ограничений; 2) Ньютона – Рафсона [4, 5], в котором на каждой итерации система нелинейных уравнений заменяется на систему линейных уравнений; 3) численного метода для решения системы линейных алгебраических уравнений.

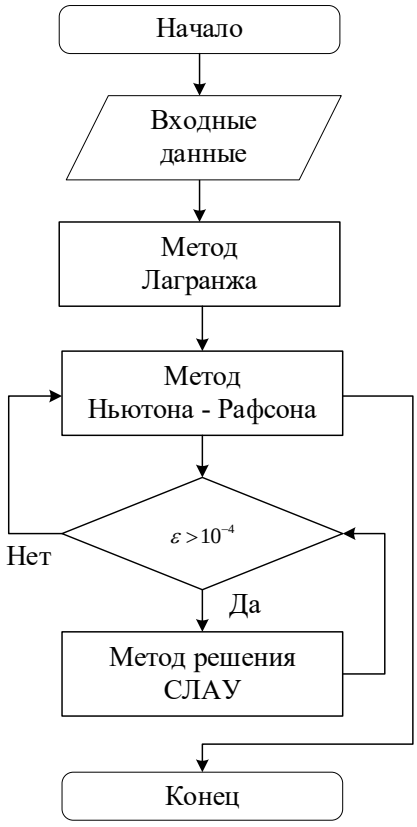


Рис. 1. Порядок работы численных методов в задаче нахождения равновесного состава сложной многокомпонентной системы

Проблемой является большое время вычисления равновесного состава сложной многокомпонентной системы при программной реализации принципа максимума энтропии. Наиболее трудоёмкой частью этой задачи является решение системы линейных алгебраических уравнений (СЛАУ). Для решения СЛАУ применяют как прямые, так и итерационные методы. СЛАУ, получаемая в ходе выполнения, не решается итерационными методами Гаусса – Зайделя и методом Якоби из-за требования к матрице быть положительно определенной. Задача нахождения равновесного состава решалась прямыми методами Гаусса [6] и LUP-разложения, при этом при количестве неизвестных более 150 скорость нахождения решения

при реализации метода LUP-разложения увеличилась в 3.6 – 4.4 раза [7]. Практические результаты авторов при решении СЛАУ показали, что методы подпространства Крылова (метод бисопряженных градиентов и квадратичный метод сопряженных градиентов) показывают наилучшие результаты по скорости нахождения решения для больших разреженных матриц [8, 9].

Сравним скорости расчета градиентными методами (квадратичный метод сопряженных градиентов и метод стабилизированных бисопряженных градиентов), показывающие быстрое время расчета для разреженных матриц [10], с методами Гаусса и LUP-разложения, применяемыми только для квадратичных матриц.

2. Методы решения систем линейных алгебраических уравнений

2.1. Квадратичный метод сопряженных градиентов (CGS – Conjugate Gradient Squared)

Алгоритм квадратичного метода сопряженных градиентов [11, 12] разработан на основе метода бисопряженных градиентов (BiCG) [13]. В алгоритме отсутствует умножение на транспонированную матрицу, поэтому алгоритм CGS работает быстрее, чем алгоритм BiCG [13].

Алгоритм квадратичного метода сопряженных градиентов имеет вид:

1. Задать начальное приближение x_0 (обычно $x_0 = 0$) и вычислить $r_0 = b - Ax_0$.
2. Выбрать вектор начальной невязки $\bar{r}$, удовлетворяющий условию $(r_0, \bar{r}) \neq 0$.
3. Цикл по переменной i = 1, 2, ... Max_N до сходимости метода.
4. $\rho_{i-1} = (\bar{r}, r_{i-1})$.
5. Если $\rho_{i-1} = 0$, система не решается.
6. Если i = 1,

$$u_1 = r_0$$
$$p_1 = u_1$$

7. Иначе

$$\beta_{i-1} = \left(\rho_{i-1} / \rho_{i-2}\right)$$
$$u_i = r_i + \beta_{i-1} \cdot q_{i-1}$$
$$p_i = u_i + \beta_{i-1}\left(q_{i-1} + \beta_{i-1} p_{i-1}\right)$$

8. Найти $\bar{p}$ из системы $M\bar{p} = p_i$:

$$\bar{v} = A\bar{p}$$
$$\alpha_i = \rho_{i-1} / \left(\bar{r}, \bar{v}\right).$$
$$q_i = u_i - \alpha_i \bar{v}$$

9. Найти $\bar{u}$ из системы $M\bar{u} = u_i + q_i$:

$$x_i = x_{i-1} + \alpha_i \bar{u}$$
$$\bar{q} = A\bar{u}$$
$$r_i = r_{i-1} + \alpha_i \bar{q}$$

10. Если $\|r\|_2 / \|r_0\|_2 < \theta$, то x_i решение найдено, выход из цикла.
11. Увеличить i.

Конец цикла.

Для реализации рассмотренного метода используется предобуславливатель неполного LU-разложения. Данный предобуславливатель превосходит по точности и скорости сходимости методы с предобуславливателем Якоби [14].

Для квадратичного метода сопряженных градиентов алгоритмическая сложность определяется по формуле [10]:

$$f_{SGC}(N) = 10N^2 + 16N + 1 + N_{it}(20N^2 + 43N + 15) + (N_{it} - 1)(10N + 4), \quad (1)$$

где N – количество неизвестных, N_{it} – количество итераций.

2.2. Метод стабилизированных бисопряженных градиентов

Метод стабилизированных бисопряженных градиентов (BiCGStab) был предложен в 1992 г. Х. ван дер Ворстом [15]. Этот метод является улучшенной версией метода бисопряженных градиентов, разработанного Флетчером в 1975 году. Метод бисопряженных градиентов обладает неустойчивостью, итерационный процесс может оборваться без возможности его продления [16]. Для больших систем линейных алгебраических уравнений он показал свою эффективность в скорости решения [8]. В статье [8] скорость нахождения решения методом BiCGStab оказалась выше, чем у метода Зейделя и метода верхней релаксации примерно в 10 – 20 раз при размерности системы 809 элементов, а время расчета уменьшилось в 2.5 – 6 раз. Для повышения эффективности метода может применятся предобусловливание.

Решение СЛАУ в матричной форме состоит в нахождении неизвестных переменных:

$$Ax = b, \quad (2)$$

где $A = (a_{ij})$ – квадратная матрица коэффициентов, $b = (b_1, b_2, \ldots b_N)$ – вектор свободных членов, N – размерность матрицы.

Алгоритм сопряженного бистабилизированного метода представим в следующем виде:

1. Задать начальные значения x_0 и вычислить $r_0 = b - Ax_0$.
2. Выбрать вектор начальной невязки, удовлетворяющий условию $(r_0, \bar{r}) \neq 0$.
3. Цикл по переменной i = 1,2 ... Max_N.
4. $\rho_{i-1} = (\bar{r}, r_{i-1})$.
5. Если $\rho_{i-1} = 0$, система не решается.
6. Если i = 1,

$p_i = r_{i-1}$.

7. Иначе

$$\beta_{i-1} = (\rho_{i-1} / \rho_{i-2}) / (\alpha_{i-1} / \omega_{i-1})$$
$$p_i = r_{i-1} + \beta_{i-1}(p_{i-1} - \omega_{i-1} v_{i-1})$$

8. Найти $\bar{p}$ из системы $M\bar{p} = p_i$:

$$v_i = A\bar{p}$$
$$\alpha_i = \rho_{i-1} / (\bar{r}, v_i).$$
$$s = r_{i-1} - \alpha_i v_i$$

9. Если $\|s\|_2 / \|r_0\|_2 < \theta$, то $x_i = x_{i-1} + \alpha_i \bar{p}$ решение найдено, выход из цикла.
10. Найти $\bar{s}$ из системы $M\bar{s} = s$:

$$t = A\bar{s}$$
$$\omega_i = (t, s) / (t, t)$$
$$x_i = x_{i-1} + \alpha_i \bar{p} + \omega_j \bar{s}$$
$$r_i = s - \omega_i t$$

11. Если $\|s\|_2 / \|r_0\|_2 < \theta$, то x_i решение найдено, выход из цикла.

12. Увеличить i.

Конец цикла.

Алгоритмическая сложность BiCGStab [10] может быть вычислена по следующим формулам в зависимости от ветки алгоритма:

$$f_{BiCGStab}(N) = 5N^2 + 13N + 3 + N_{it}(10N^2 + 31N + 16) + (N_{it} - 1)(10N^2 + 33N + 11), \quad (3)$$

$$f_{BiCGStab}(N) = 5N^2 + 8N + 2 + N_{it}(20N^2 + 64N + 37). \quad (4)$$

Как видно из рис. 1, всегда будет применяться численный метод Лагранжа, затем на каждой итерации метода Ньютона – Рафсона будет вызываться численный метод для решения СЛАУ. Замеряться будет общее время расчета для нахождения равновесного состава.

3. Вычислительный эксперимент

Тестирование программы, реализующей численные методы, производилось на операционной системе Microsoft Windows 10 с использованием центрального процессора AMD Phenom 955 BE, 12 Gb DDR3 оперативной памяти. Вычисления производились с использованием одного ядра процессора. Программный код написан на языке C# в среде Visual Studio 2022.

3.1. Сравнение программных реализаций численных методов решения СЛАУ

В численном эксперименте сравнивалась работоспособность следующих методов: 1) прямого метода Гаусса; 2) разложения LUP; 3) квадратичного метода сопряженных градиентов и 4) метода бисопряженных градиентов (рис. 2).

Так как СЛАУ малой размерности решаются за несколько миллисекунд, а время выполнения одной и той же задачи может отличаться вплоть до двух десятков миллисекунд, замеряли серию вычислений для выбранного метода. В табл. 1 приведено сравнение быстродействия методов Гаусса, CGS, BiSGStab и LUP по общему времени измерения.

Таблица 1. Общее время нахождения решения в зависимости от метода при вычислении на одном ядре

Размерность матрицы	Шаг по температуре	Количество шагов	Общее время измерения, мс			
	ΔT	m	Гаусса	CGS	BiSGStab	LUP
15	1	1001	4467	5299	4847	3852
79	50	21	7163	5358	5143	2943
153	100	11	24810	16846	16248	7151
174	100	11	33873	22491	21820	9079
214	100	11	73134	46063	45129	18160
355	200	6	194035	114241	108355	39795
376	200	6	220054	132257	128794	49222

В табл. 2 приведено сравнение быстродействия методов CGS, BiSGStab и LUP с методом Гаусса.

Таблица 2. Сравнение средних скоростей вычисления
при использовании различных методов решения СЛАУ на одном ядре

Размерность матрицы	Ускорение метода по сравнению с методом Гаусса		
	τГ / τCGS	τГ / τBiCSGStab	τГ / τLUP
15	0.84	0.92	1.16
79	1.34	1.39	2.43
153	1.47	1.53	3.47
174	1.51	1.55	3.73
214	1.59	1.62	4.03
355	1.70	1.79	4.88
376	1.66	1.71	4.47

Стоит отметить, что матрица, подаваемая на вход при реализации метода решения СЛАУ, не является разреженной, поэтому решалась полная система линейных алгебраических уравнений. В работе [10] для решения разреженной матрицы получилось ускорение от 1.2 до 4.4 раза при использовании стабилизированного метода бисопряженных градиентов (BiCGStab) относительно метода Гаусса. Результаты вычислительных экспериментов приведены на рис. 2.

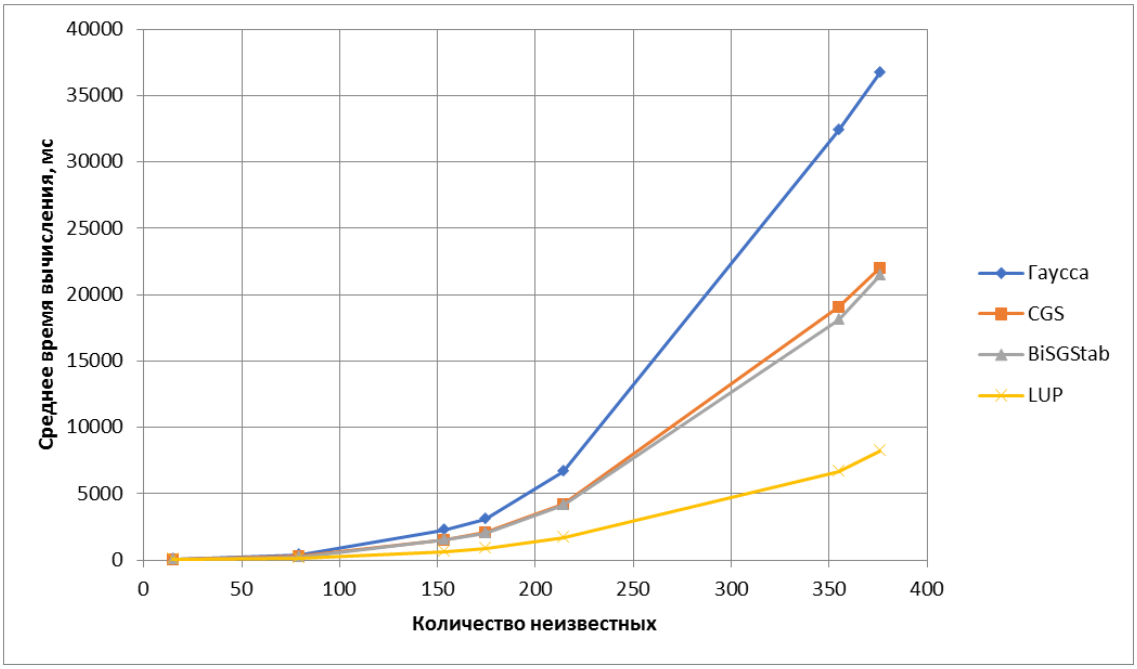


Рис. 2. Сравнение времени поиска решения методами Гаусса, CGS, BiCGStab, LUP

Как видно из табл. 1 и 2, самым быстрым вляется алгоритм метода LUP-разложения: он оказалсяся быстрее остальных при размерности матрицы свыше 150 элементов в 3.60 – 4.55 раза. В методах CGS и BiCGStab наиболее трудоемкой задачей является реализация метода предобуславливания матрицы, в данном случае использовался метод ILU(0) – неполного разложения матрицы.

Так как за основу для алгоритмов CGS и BiCGStab были взяты алгоритмы, для которых алгоритмическая сложность вычисляется по формулам (1), (3) и (4), то сравним полученные результаты с теоретической сложностью этих алгоритмов, как показано в табл. 3.

Как видно из табл. 3, относительная ошибка между отношениями скоростей работы методов сопряженного градиента и стабилизированных бисопряженных градиентов не превышает 6.45 %.

Таблица 3. Сравнение теоретического и практического времени вычисления для методов CGS и BiCGStab

Размерность матрицы, N	Количество итераций N_{it}	Практическое ускорение	Теоретическое ускорение	Относительная ошибка, %
		τCGS/τBiCGStab	τCGS/τBiCGStab	
15	35	1.044	0.979	6.45
79	37	1.032	1.006	2.51
153	39	0.999	1.009	1.03
174	37	1.026	1.010	1.52
214	50	0.986	1.007	2.14
355	42	0.997	1.010	1.34
376	42	0.992	1.010	1.79

В табл. 4, как и в табл. 1, приведено сравнение скоростей нахождения равновесного состава с использованием методов Гаусса, CGS, BiSGStab и LUP. Результаты были получены при параллельном запуске 4 потоков, которые нагрузят все 4 ядра центрального процессора.

Таблица 4. Общее время нахождения решения для различных методов при параллельной реализации на 4 ядрах

Размерность матрицы	Шаг по температуре	Количество шагов	Общее время измерения, мс			
	ΔT	m	Гаусса$_4$	CGS$_4$	BiSGStab$_4$	LUP$_4$
15	1	1001	1554	1793	1717	1287
79	50	21	2430	2374	2301	959
153	100	11	9083	6486	6495	2118
174	100	11	12170	8602	8387	2781
214	100	11	25096	17217	17462	5499
355	200	6	63319	40573	40705	13919
376	200	6	81648	49604	49986	17756

Увеличение средних скоростей вычисления методами CGS, BiCGStab и LUP относительно метода Гаусса при параллельной реализации алгоритма показано в табл. 5.

Таблица 5. Сравнение средних скоростей вычисления с использованием различных методов решения СЛАУ на 4 ядрах

Размерность матрицы	Ускорение метода по сравнению с методом Гаусса		
	τГ$_4$ / τCGS$_4$	τГ$_4$ / τBiCSGStab$_4$	τГ$_4$ / τLUP$_4$
15	0.87	0.91	1.21
79	1.02	1.06	2.53
153	1.40	1.40	4.29
174	1.41	1.45	4.38
214	1.46	1.44	4.56
355	1.56	1.56	4.55
376	1.65	1.63	4.60

В табл. 6 приведены значения среднего увеличения скорости вычислений при использовании 4 ядер процессора по сравнению с использованием 1 ядра.

Таблица 6. Сравнение скорости вычисления при использовании 4 ядер и 1 ядра при реализации различных методов решения СЛАУ

Размерность матрицы	Отношение времен выполнения при вычислении в одном и четырех ядрах			
	$\tau Г/ \tau Г_4$	$\tau CGS/ \tau CGS_4$	$\tau BiCSGStab / \tau BiCSGStab_4$	$\tau LUP/ \tau LUP_4$
15	2.87	2.96	2.82	2.99
79	2.95	2.26	2.24	3.07
153	2.73	2.60	2.50	3.38
174	2.78	2.61	2.60	3.26
214	2.91	2.68	2.58	3.30
355	3.06	2.82	2.66	2.86
376	2.70	2.67	2.58	2.77

Ранее в статье [17] было получено максимальное ускорение работы методов от 2.5 до 3.22 раза. Из табл. 5 видно, что ускорение при 4 ядрах составило от 2.24 до 3.38 раз. Самым быстрым оказался алгоритм метода LUP при использовании 4 ядер. При этом общее время расчета равновесного состава для шести разных температур методом LUP сократилось с 49.2 до 17.8 секунд.

3.2. Сравнение быстродействия алгоритмов численных методов решения СЛАУ с использованием динамически подключаемых библиотек линейной алгебры

Поскольку основной программный код был написан на языке C# в среде Visual Studio 2022, то рассмотрим динамически подключаемые библиотеки, находящиеся в свободном доступе: 1) Alglib (скачиваемый пакет alglibnet2 содержит один файл alglibnet2.dll) 257 тыс. скачиваний; 2) ILNumerics (пробная версия на 14 дней, для запуска требуется регистрация пробной лицензии, 5 файлов ILNumerics.Computing, ILNumerics.Core, ILNumerics.CorePlatforms, ILNumerics.Core.Runtime, ILNUmeritcs.F2NET.LAPACK) 33 тыс. скачиваний; 3) MathNet.Numerics (название скачиваемой библиотеки MathNet.Numerics.dll для основной работы, а также для использования ускорения от Intel MKL – две дополнительные библиотеки: libMathNetNumericsMKL.dll и libiomp5md.dll) 29.3 млн. скачиваний; 4) Accord.Math (скачиваются три библиотеки Accord.dll, Accord.Math.Core.dll и Accord.Math.dll) 5.16 млн. скачиваний.

В библиотеке Alglib поддержка ускорения от Intel MKL присутствует, но дополнительные файлы, к сожалению, не были отправлены после запроса о регистрации. В динамически подключаемых библиотеках Alglib и Accord все реализованные методы вызываются вручную, а в библиотеках ILNumerics и MathNet используется один вызов метода для решения СЛАУ, который в автоматическом режиме выбирает походящую реализацию в зависимости от типа входной матрицы.

В задаче нахождения равновесного состава термодинамической системы исходная матрица для решения СЛАУ является симметричной, но не положительно определенной. Поэтому разложение Холецкого в динамически подключаемых библиотеках ALGLIB, ILNumerics и Accord не выполняется. В табл. 7 приведем сравнение быстродействия методов при параллельной реализации на 4 ядрах, реализованных в пакете Alglib. При этом сравнивалось, как и в разделе 3.1, общее время решения задачи, а не время решения СЛАУ. В данной библиотеке есть реализация задания матриц с вещественными числами (r) и комплексными числами (c).

Таблица 7. Общее время работы программы в зависимости от метода в библиотеке ALGLIB при параллельной реализации на 4 ядрах

Размерность матрицы	Шаг по температуре	Количество шагов	Общее время измерения, мс					
	ΔT	m	rГаусса$_4$	cГаусса$_4$	rLU$_4$	cLU$_4$	rmix$_4$	cmix$_4$
15	1	1001	6798	6299	2462	5393	3310	7065
79	50	21	10369	3969	1560	3777	1588	3941
153	100	11	34202	8802	2426	7770	2795	9479
174	100	11	54384	11433	3129	10339	3581	13099
214	100	11	98447	22515	6022	20181	6846	26052
355	200	6	245372	55261	13802	51059	15951	59590
376	200	6			17321			

Из табл. 7 видно, что самым быстрым оказался алгоритм метода LU с вещественными числами, который по скорости вычисления близок к методу LUP (табл. 4), реализованному авторами.

В табл. 8 приведены результаты скорости нахождения решения с использованием динамически подключаемых библиотек ILNumerics, MathNet и Accord. В первых двух библиотеках алгоритмы для решения задачи выбирались автоматически, а в библиотеке Accord использовался метод LU-разложения.

Таблица 8. Общее время выполнения программы в зависимости от метода в библиотеках ILNumerics, MathNet и Accord при параллельной реализации на 4 ядрах

Размерность матрицы	Шаг по температуре	Количество шагов	Общее время измерения, мс			
	ΔT	m	ILNumerics$_4$	MathNet$_4$	Accord$_4$	MathNet$_4$ + MKL
15	1	1001	28030	1265	1104	1056
79	50	21	2689	1077	492	480
53	100	11	2619	1207	996	895
174	100	11	3461	1301	1219	1030
214	100	11	5754	2418	2106	1578
355	200	6	12370	5230	4238	2562
376	200	6	17803	5958	5186	2977

Как видно из табл. 8 самыми быстрыми оказались алгоритмы методов, реализованных в библиотеке MathNet с поддержкой MKL (для поддержки требуются файлы библиотек 155 Мб и 95.5 Мб для 64-битной и 32-битной версий программы соответственно). Реализация авторов сравнима по скорости выполнения с библиотекой ILNumerics и проигрывает по скорости вычисления с использованием библиотеки MathNet + MKL при решении задачи для более 150 элементов от 2.6 до 6 раз. При возрастании размерности матриц отрыв в скорости вычисления с использованием библиотек MathNet + MKL будет увеличиваться.

4. Заключение

Программно реализованные авторами градиентные методы CGS и BiCGStab показали близкие по скорости вычислений результаты с небольшим преимуществом метода стабилизированного бисопряженного градиента. Ускорение работы градиентных методов по сравнению с методом Гаусса составило 1.42 – 1.71 при размерности матрицы свыше 150 элементов. Программа с использованием метода LUP показала лучшие результаты: ускорение относительно метода Гаусса составило 4.29 – 4.6 раза. Параллельная реализация алгоритмов на 4-ядерном процессоре дала ускорение в вычислении от 2.24 до 3.38 раз для всех методов.

Сравнение скорости расчетов при использовании динамически подключаемых библиотек линейной алгебры, находящихся в открытом доступе, показало, что самый быстрый – реализованный авторами алгоритм метода LUP.

Литература

1. *Сеченов П. А., Рыбенко И. А.* Численный метод и математическая модель нахождения равновесного состава термодинамической системы программного комплекса T-Energy // Вестник Дагестанского государственного технического университета. Технические науки. 2022. Т. 49, № 4. С. 104–112.
2. *Базилевский М. П., Власенко Л. Н.* Оценивание моделей парной линейной регрессии с параметрами в виде матриц линейных операторов двумерного векторного пространства // Моделирование, оптимизация и информационные технологии. 2020. Т. 8, № 1 (28).
3. *Андрианов И. К.* Численная модель многокритериальной оптимизации тепловой защиты оболочечных элементов в условиях теплового и силового нагружения // Ученые записки Комсомольского-на-Амуре государственного технического университета. 2021. Т. 51, № 3. С. 14–20.
4. *Покаместов Д. А., Кологривов В. А., Бабур-Каратели Г. П., Крюков Я. В.* Расчёт зон обслуживания кластера базовых станций систем сотовой связи при заданном распределении абонентов // Вестник СибГУТИ. 2017. Т. 38, № 2. С. 26–34.
5. *Журтов А. В., Хежев Т. А., Хаджишалапов Г. Н., Карданов А. А., Шогенцуков М. А.* Напряженно-деформированное состояние двухслойных армоцементных оболочек в условиях пожара // Вестник Дагестанского государственного технического университета. Технические науки. 2022. Т. 49, № 1. С. 112–121.
6. *Белов В. М., Зырянова Е. В., Рябова Е. В.* Интервально-статистический метод определения параметров функций принадлежности в задачах оценки качества систем комплексной безопасности // Вестник СибГУТИ. 2016. Т. 34, № 2. С. 85–96.
7. *Сеченов П. А.* Сравнение быстродействия численных методов Гаусса и LUP-разложения в задаче нахождения равновесного химического состава // Вестник Воронежского государственного технического университета. 2023. Т. 19, № 2. С. 79–85.
8. *Пузикова В. В.* Решение систем линейных алгебраических уравнений методом BiCG-STAB с предобусловливанием // Вестник Московского государственного технического университета им. Н. Э. Баумана. Серия Естественные науки. 2011. № S1. С. 124–133.
9. *Чадов С. Н.* О решении разреженных систем линейных уравнений при помощи стабилизированного метода бисопряженных градиентов // Вестник Ивановского государственного энергетического университета. 2007. № 3. С. 86–88.
10. *Ахунов Р. Р.* Алгоритмы и комплекс программ для итерационного решения систем линейных алгебраических уравнений при анализе полосковых структур методом моментов: дис. … канд. тех. наук. Томск. 2018. 199 с.
11. *Серышева И. А., Хрусталев Ю. П.* Формализованная методика обработки измерительной информации, получаемой в процессе функционирования группового эталона времени и

частоты // Информационные и математические технологии в науке и управлении. 2018. Т. 12, № 4. С. 163–172.
12. *Сороковиков П. С., Горнов А. Ю.* Пакет программ meopt для решения невыпуклых задач параметрической идентификации // Информационные и математические технологии в науке и управлении. 2022. Т. 26, № 2. С. 53–60.
13. *Горбаченко В. И.* Вычислительная линейная алгебра с примерами на MATLAB. СПб.: БХВ-Петербург, 2011. 320 с.
14. *Гоголева С. Ю.* Предобусловливание на основе LU-разложения в итерационных методах для решения систем линейных алгебраических уравнений с разреженными матрицами // Материалы IV Международной конференция и молодёжной школы «Информационные технологии и нанотехнологии», Самара, 24–27 апреля, 2018. С. 1493–1498.
15. *Vander Vorst H. A.* Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for solution of non-symmetric linear systems // SIAM J. Sci. Stat. Comp. 1992. № 2. P. 631–644.
16. *Куксенко С. П., Газизов Т. Р.* Итерационные методы решения системы линейных алгебраических уравнений с плотной матрицей. Томск: Томский государственный университет, 2007. 208 с.
17. *Сеченов П. А., Кожемяченко В. И., Рыбенко И. А.* Параллельная реализация алгоритма расчёта равновесного состава в программном комплексе T-Energy // Вестник Поволжского государственного технологического университета. Серия: Радиотехнические и инфокоммуникационные системы. 2023. Т. 57, № 1. С. 64–73.

Сеченов Павел Александрович

к.т.н., доцент кафедры прикладных информационных технологий и программирования, Сибирский государственный индустриальный университет (СибГИУ, 654007, Новокузнецк, ул. Кирова, зд. 42), тел. +7 3843 701 563, e-mail: pavesa89@mail.ru, ORCID ID: 0000-0003-4551-3822.

Рыбенко Инна Анатольевна

д.т.н., заведующий кафедрой прикладных информационных технологий и программирования, Сибирский государственный индустриальный университет (СибГИУ, 654007, Новокузнецк, ул. Кирова, зд. 42**)**, тел. +7 3843 701 563, e-mail: rybenkoi@mail.ru, ORCID ID: 0000-0003-1679-0839.

Авторы прочитали и одобрили окончательный вариант рукописи.

Авторы заявляют об отсутствии конфликта интересов.

Вклад соавторов: *Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.*

Comparison of Software Implementations of SLAE Solution Methods in the Problem of Finding the Equilibrium Composition of a Complex Multicomponent Heterogeneous System

Pavel A. Sechenov, Inna A. Rybenko

Siberian State Industrial University, Novokuznetsk

Abstract: The article presents a comparative analysis of the software implementation performance of numerical methods in solving the problem of finding the equilibrium composition of a complex multicomponent heterogeneous system. The task of finding the equilibrium composition of the system is divided into the following subtasks: 1) taking into account constraints (Lagrange method); 2) finding the maximum function of a nonlinear function: 2.1) converting a function into a system of linear equations (Newton Raphson method); 2.2) using numerical methods to solve a system of linear algebraic equations. An analytical review of the literature data has shown that gradient methods have better performance when solving systems of linear algebraic equations. Therefore, the article compared the performance of the software implementation of the entire algorithm using direct (Gauss, LUP decomposition) and iterative methods (conjugate gradient method, biconjugate gradient stabilized method) for solving a system of linear algebraic equations. The calculation speed of the developed program was also compared using dynamically connected libraries Alglib, ILNumerics, MathNet, Accord to solve the SLAE in the problem of finding the equilibrium composition of a thermodynamic system.

Keywords: numerical methods, conjugate gradient method, method of stabilized bisjoint gradients, speed, equilibrium composition, dynamic link library.

For citation: Sechenov P. A., Rybenko I. A. Comparison of software implementations of SLAE solution methods in the problem of finding the equilibrium composition of a complex multicomponent heterogeneous system (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 3, pp. 86-98. https://doi.org/10.55648/1998-6920-2024-18-3-86-98.

Content is available under the license Creative Commons Attribution 4.0 License

© Sechenov P. A., Rybenko I. A., 2024

The article was submitted: 11.01.2024;
revised version: 25.02.2024;
accepted for publication 02.04.2024.

References

1. Sechenov P. A., Rybenko I. A. CHislennyj metod i matematicheskaya model' nahozhdeniya ravnovesnogo sostava termodinamicheskoj sistemy programmnogo kompleksa T-Energy [Numerical method and mathematical model for finding the equilibrium composition of the thermodynamic system of the T-Energy software package]. *Herald of Dagestan State Technical University. Technical Sciences*, 2022, vol. 49, no. 4, pp. 104-112.
2. Bazilevskij M. P., Vlasenko L. N. Ocenivanie modelej parnoj linejnoj regressii s parametrami v vide matric linejnyh operatorov dvumernogo vektornogo prostranstva [Estimation of pair linear regression models with parameters in the form of linear operator matrices of two-dimensional vector space]. *Modeling, Optimization and Information Technology*, 2020, vol. 8. no 1 (28).
3. Andrianov I. K. CHislennaya model' mnogokriterial'noj optimizacii teplovoj zashchity obolochechnyh elementov v usloviyah teplovogo i silovogo nagruzheniya [Numerical multi-criteria optimization model of the shell thermal protection under conditions of thermal and force loading]. *Scholarly Notes of KNASTU*, 2021, vol. 51, no. 3, pp. 14-20.
4. Pokamestov D. A., Kologrivov V. A., Babur-Karatelli G. P., Kryukov YA. V. Raschyot zon obsluzhivaniya klastera bazovyh stancij sistem sotovoj svyazi pri zadannom raspredelenii abonentov [Calculation of base station coverage cluster for a given subscriber distribution]. *Vestnik SibGUTI*, 2017, no. 2, pp. 26-34.
5. Zhurtov A. V., Khezhev T. A., Khadzhishalapov G. N., Kardanov A. A., Shogentsukov M. A. Napryazhenno-deformirovannoe sostoyanie dvuhslojnyh armocementnyh obolochek v usloviyah pozhara [Stress strain state of double-layered ferrocement shells under the fire condition]. *Herald of Dagestan State Technical University. Technical Sciences*, 2022, no. 49(1), pp. 112-121.

6. Belov V. M., Zyrjanova E. V., Rjabova E. V. Interval'no-statisticheskij metod opredelenija parametrov funkcij prinadlezhnosti v zadachah ocenki kachestva sistem kompleksnoj bezopasnosti [Interval-statistical method of membership functions parameters determination in problems of the integrated security system quality assessment]. *Vestnik SibGUTI*. 2016, no. 2, pp. 85-96.
7. Sechenov P. A. Sravnenie bystrodejstviya chislennyh metodov Gaussa i LUP-razlozheniya v zadache nahozhdeniya ravnovesnogo himicheskogo sostava [Comparison of the speed of numerical methods of Gauss and LUP decomposition in the problem of finding the equilibrium chemical composition]. *The Bulletin of Voronezh state technical university*, 2023, vol. 19. no. 2, pp. 79-85.
8. Puzikova V. V. Reshenie sistem linejnyh algebraicheskih uravnenij metodom BiCGSTAB s predobuslovlivaniem [Solution of systems of linear algebraic equations by BiCGStab method with preconditioning]. *Herald of the Bauman Moscow State Technical University. Natural Sciences Series*, 2011, no. S1, pp. 124-133.
9. CHadov S. N. O reshenii razrezhennyh sistem linejnyh uravnenij pri pomoshchi stabilizirovannogo metoda bisopryazhennyh gradientov [On the solution of sparse systems of linear equations using the stabilized method of bisjoint gradients]. *Vestnik of the Ivanovo State Power Engineering University*, 2007, no. 3, pp. 86-88.
10. Ahunov R. R. *Algoritmy i kompleks programm dlya iteracionnogo resheniya sistem linejnyh algebraicheskih uravnenij pri analize poloskovyh struktur metodom momentov* [Algorithms and software package for iterative solution of systems of linear algebraic equations in the analysis of strip structures by the method of moments]. Ph.D. thesis, Tomsk. 2018, 199 p.
11. Serysheva I. A., Hrustalev YU. P. Formalizovannaya metodika obrabotki izmeritel'noj informacii, poluchaemoj v processe funkcionirovaniya gruppovogo etalona vremeni i chastity [The formalized technique of processing measuring information obtained during the operation of a group standard of time and frequency]. *Information and mathematical technologies in science and management*, 2018, no. 4 (12), pp. 163-172.
12. Sorokovikov P. S., Gornov A. YU. Paket programm meopt dlya resheniya nevypuklyh zadach parametricheskoj identifikacii [MEOPT software package for solving non-convex problems of parametric identification]. *Information and mathematical technologies in science and management*, 2022, no. 2 (26). pp. 53-60.
13. Gorbachenko V. I. *Vychislitel'naya linejnaya algebra s primerami na MATLAB* [Computational linear algebra with MATLAB examples Saint Petersburg, BHV-Peterburg, 2011, 320 p.
14. Gogoleva S. YU. Predobuslovlivanie na osnove LU-razlozheniya v iteracionnyh metodah dlya resheniya sistem linejnyh algebraicheskih uravnenij s razrezhennymi matricami [LU-based preconditioning in iterative methods for solving systems of linear algebraic equations with sparse matrices] *IV Mezhdunarodnaya konferenciya i molodyozhnaya shkola "Informacionnye tekhnologii i nanotekhnologii"*, Samara, 2018, pp. 1493-1498.
15. Vander Vorst H. A. Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for solution of non-symmetric linear systems. *SIAM J. Sci. Stat. Comp*, 1992, no. 2. pp. 631-644.
16. Kuksenko S. P., Gazizov T. R. *Iteracionnye metody resheniya sistemy linejnyh algebraicheskih uravnenij s plotnoj matricej* [Iterative methods for solving a system of linear algebraic equations with a dense matrix]. Tomsk: Tomskij gosudarstvennyj universitet, 2007, 208 p.
17. Sechenov P. A., Kozhemyachenko V. I., Rybenko I. A. Parallel'naya realizaciya algoritma raschyota ravnovesnogo sostava v programmnom komplekse T-Energy [Parallel Implementation of the Equilibrium Composition Calculation Algorithm in the T-Energy Software Package]. *Vestnik of Volga State University of Technology. Ser.: Radio Engineering and Infocommunication Systems*, 2023, vol. 57, no. 1, pp. 64-73.

Pavel A. Sechenov

Ph.D. (Tech.), associate professor at the Department of Applied Information Technologies and Programming, Siberian State Industrial University (SibSIU, Russia, 654007, Novokuznetsk, Kirov St. 42), e-mail: pavesa89@mail.ru, ORCID: 0000-0003-4551-3822.

Inna A. Rybenko

D. Sc. (Tech.), Assoc, Head of the Department of Applied Information Technologies and Programming, Siberian State Industrial University (SibSIU, Russia, 654007, Novokuznetsk, Kirov St. 42), e-mail: rybenkoi@mail.ru, ORCID ID: 0000-0003-1679-0839.

DOI: 10.55648/1998-6920-2024-18-3-99-114 УДК 519.862.2

Финансовый баланс экономики региона

Т. Л. Самков

Сибирский гос. унив. телекоммуникаций и информатики (СибГУТИ)
Новосибирский государственный технический университет (НГТУ)

Аннотация: Статья продолжает серию работ, описывающих модель АТМО (анализа территориальных мультисекторных объектов). Она отражает структуру и функциональность финансового блока этой модели, относящейся к Леонтьевским моделям «затраты – выпуск». Блок схематизирует взаиморасчеты производственных и сбытовых предприятий в рамках взаимных поставок выпускаемой и ввозимой продукции внутри региона. Данные денежные отношения включают учет суммарной себестоимости производства и распространения, ввоза и сбыта продукции. Они соответствуют производственными и сбытовыми компаниями региона для всех его секторов с последующим соотнесением с полученной выручкой от производственной и сбытовой деятельности. Кроме учета сбытовой составляющей получающейся операционной прибыли сектора, новизна модели состоит в учете субсидий и госфинансирования, отражении затрат на продукцию, произведенную вне региона. Дополнительно введены механизмы учета кредитных и налоговых выплат на общий финансовый итог работы секторов. Разработаны формальные механизмы взаимосвязи инвестиций в логистику и затрат на ввоз и реализацию внерегиональных товаров и оборудования. Модель включает в себя соотношения, задающие движение к оптимальным значениям тех ее показателей, которые, основываясь на принципах промышленного развития, предусматривают наличие средств на покупку дополнительных станков и оборудования, увеличивающих объем производства и сбытовых операций. Эти средства имеются в наличии у сектора, если его операционная прибыль больше нуля, и они отсутствуют в противном случае.

Ключевые слова: модель «затраты – выпуск», логистические коэффициенты, сектор, сегмент, удельные затраты, себестоимость, доля логистического перерасхода/экономии.

Для цитирования: Самков Т. Л. Финансовый баланс экономики региона // Вестник СибГУТИ. 2024. Т. 18, № 3. С. 99–114. https://doi.org/10.55648/1998-6920-2024-18-3-99-114.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Самков Т. Л., 2024

Статья поступила в редакцию 06.04.2024;
переработанный вариант – 02.08.2024;
принята к публикации 09.08.2024.

1. Введение

Данная работа продолжает цикл статей, посвященных описанию разработанной автором модели АТМО – модели анализа территориальных мультисекторных объектов.

Ранее [1] был описан блок межотраслевого продукционного баланса, основообразующий для модели АТМО. Данная же работа посвящена блоку финансового межотраслевого баланса секторов региона в рамках АТМО, дополняющего продукционный баланс. Здесь исследуются потребление секторами выпущенной и ввезенной продукции на единицу произведенной сектором продукции. Это потребление выражается элементами транспонированных матриц технологических и логистических коэффициентов, дающих в сумме совокупные затраты на

сырье и материал, поставляемые другими секторами согласно этим коэффициентам. Они отражают структуры основных затрат, но в блоке имеется и ряд других статей расходов.

Структуру денежных потоков сектора формирует набор следующих факторов:

- хеджирование рисков предприятиями соответствующих секторов;
- ценовая политика предприятий-производителей;
- стимулирование вывоза продукции, выпущенной в регионе;
- ценовая политика торговых предприятий;
- политика стимулирования роста через транзит продукции;
- технологические пропорции затрат продукции секторов на выпуск;
- логистические пропорции затрат продукции секторов на ввоз;
- уровень доходов работающих в секторах;
- налоговая и кредитная политика в регионах.

2. Существующие модели движения межотраслевых денежных потоков

Исследования, посвященные моделированию финансовых межотраслевых потоков, характеризуются, как правило, некоторой фрагментарностью и узкой направленностью.

В [2] на основе классической модели «затраты – выпуск» выявляются взаимосвязи между эндогенными и экзогенными счетами, для чего использована матрица финансовых потоков (МФП). Она получается транспонированием матрицы технологических коэффициентов с получением на ее основе матрицы средней склонности к потреблению. Для этого производится деление каждого показателя в эндогенном столбце на сумму показателей соответствующего столбца. Доходы и расходы в рамках модели выражены в параметрах коэффициентов матрицы средней склонности к потреблению и вектора доходов от выпуска продукции.

На основе классической схемы круговых потоков, основанных на межотраслевых взаимодействиях, в [3] предложена схема круговых потоков региона, представленная на рис. 1.

Так как социально-экономическая система региона является открытой, то его структура собственности в значительной мере определяет потоки товаров, услуг и финансовых средств.

Из схемы следует необходимость включения в анализ закономерностей развития региона всех указанных элементов и факторов с определением диапазонов изменения их значений.

Рис. 1. Схема круговых потоков региона

В [4] для системы нескольких отраслей использованы следующие показатели:

- валовые выпуски продукта отраслями;
- объёмы поставок продукции отраслей друг другу;
- цены единиц продукции отраслей;
- валовые объемы взаимных поставок отраслей и их доходов;
- объём конечного потребления.

Введен показатель валового внутреннего продукта, формируемый на основе доходов. Он включает валовую добавленную стоимость, определяемую зависимостью ВВП от ряда показателей. Это доходы, куда входят оплата работы по найму, рентные платежи, амортизация основного капитала, доходы от финансовых активов, прибыль корпораций и доход некорпорированных предприятий. На основе этого представления строятся балансовые соотношения в виде системы линейных уравнений, где имеется следующий баланс. Доход отрасли в левой части уравновешивается потреблением товаров других отраслей (в соответствии с технологическими коэффициентами) совместно с показателями выше, формирующими ВВП, в правой части. Указанное позволяет оценить структуру экономики, проанализировать потенциальные результаты реструктуризации, создать модель изменения структуры отраслей.

Довольно часто показатели, отражающие движение финансовых средств между отраслями, используют при построении модели развития экономики региона [5]. Для этого вводятся векторы валового объема выпуска продукции и конечного спроса на нее в плановом периоде. Последний, в свою очередь, состоит из конечного потребления, накоплений и чистого экспорта. Кроме того, задействован вектор валовых объемов инвестиций, вкладываемых в увеличение производственных мощностей. В модели развития экономики региона валовой объем выпуска отрасли равен сумме затрат на продукцию других отраслей и дохода от производства продукции с учетом технологических коэффициентов. При этом валовая добавленная стоимость в отрасли состоит из зарплаты, налоговых выплат, амортизации и прибыли.

В практике применения модели МОБ разработан механизм инвестиционно-финансового перераспределения ресурсов в межотраслевом блоке производства за ряд этапов [6]. Здесь осуществлен переход от МОБ «Затраты – Выпуск» (ЗВ) к МОБ «Производство – Потребление» (ПП). В первой модели ищут валовое промежуточное потребление как стоимость продукции конкретной отрасли, расходуемой в другой отрасли, в т.ч. на поставки отечественной продукции и импорта. Также это транспортная и торгово-посредническая наценки в поставках продукции, дополненные налоговыми выплатами.

В данной модели сумма объёмов выпуска отрасли корректируется на стоимость услуг финансового посредничества с учетом условно чистой продукции. Она, в свою очередь, агрегирует затраты на возмещение затрат труда работников, чистую прибыль, а также налоги и рентные платежи. При этом показатель суммы потребности в объёмах выпуска промежуточной также корректируется с учетом услуг финансового посредничества.

В рамках перехода к модели «Потребление – Производство» вводится целевая функция на основе показателя дебетового сальдо баланса, ограниченная долей косвенно измеряемых услуг финансового посредничества.

В [7] на основе классических постановок моделей материально-вещественного состава, приведенных в базовых работах А. Г. Гранберга, показан их блочный вид. Он формируется как подробный анализ затрат на промежуточное потребление отраслями продукции не только своего района, но и из других районов. В работе также представлен фрагмент условий модели межрегионального межотраслевого финансового баланса в виде дополнительных балансов к условиям классических постановок моделей в блочном виде. Для этого используется вектор признака блока потребления домохозяйств, состоящий из ряда компонент отраслевого потребления. Это заработная плата, расходы населения на непроизводственное потребление, налоги с населения, изменение сбережений, получение трансфертов, пособий и т.д., капитальные вложения, налоги, возврат кредитов; общие налоговые доходы, сумма возвращаемых и полученных кредитов, расходы бюджетов на территории.

Процессы производства и потребления сопровождает ряд финансовых потоков. Это общий объем денежных средств из частных по районом, дополнительная эмиссия ЦБ, чистая прибыль, выплаты из региональных и федеральных бюджетов (дотации, субсидии и т.д.), кредиты, выданные банками района и ЦБ, сальдо кредиторской и дебиторской задолженностей.

Модель МОБ применима и к оценке потребительского спроса [8]. В рамках решения данной проблемы использована система функций потребительского спроса «ПАДС». Она анализирует закономерности потребления домохозяйств и является инструментом его эндогенизации в моделях межотраслевых экономики. Это дает возможность строить сценарные прогнозы расходов домохозяйств для отдельных отраслей.

Система «ПАДС» представлена балансовым уравнением, в левой части которого стоит среднедушевое потребление определенной группы товаров и услуг. В правой части оно представляется линейной комбинацией временного тренда с компонентой среднедушевых совокупных номинальных расходов/доходов. Они, в свою очередь, масштабированы ценовыми индексами для совокупных расходов с учетом инфляции и скорректированы ценовыми индексами для каждой группы товаров и услуг. А доля расходов на них учтена в совокупных номинальных расходах в базовом году.

В [9] отражена структура валовой добавленной стоимости (ВДС), возникающей в итоге производственного процесса в системе «затраты – выпуск». Авторы применили ранее разработанную межотраслевую модель разложения стоимости отечественной конечной продукции на вклады отечественного и зарубежного производителя в ее добавленной стоимости.

В модели балансируется конечное использование отечественной продукции конкретного вида в основных ценах совокупностью трех компонент. Первая – это доля ВДС в стоимости выпуска отрасли, взятой от промежуточных затрат отечественной продукции данной отрасли на единицу выпуска остальных отраслей. Вторая – доля чистых налогов на продукты в стоимости выпуска в основных ценах этой отрасли, отнесенная ко всем видам ее использования. Третья – коэффициент затрат импорта данной отрасли на выпуск единицы отечественной продукции конкретного вида, нормированный его использованием.

Данная модель не учитывает влияние инвестиционных и инновационных факторов, что корректируется использованием производственных функций, подобных функции Кобба–Дугласа. Здесь выпуск отражается в виде ряда сомножителей для конкретного вида деятельности. Это совокупная факторная производительность, среднегодовой размер фондов, численность занятых и объем инвестиций в основной капитал для инновационного/инфраструктурного сектора. При этом второй и третий сомножители возведены в степень, сформированную эластичностью выпуска по фактору капитала, а четвертый – эластичностью выпуска инновационного/инфраструктурного сектора по инвестициям в основной капитал.

В использовании финансового межотраслевого баланса имеет место исследование проектной деятельности, изменяющей объем выпуска и финансовые характеристики межотраслевой экономики региона [10] с применением двух связанных моделей:

• оптимизационная межотраслевая межрегиональная модель (ОМММ);

• финансово-экономическая модель инфраструктурного проекта (ФЭМ).

Согласно ОМММ региональные блоки балансовых ограничений по выпуску модифицированной модели изменяются в результате проектной деятельности следующим образом.

В составе модифицированной ОМММ наряду с обычными поставками продукции отдельной отрасли другим отраслям присутствуют и материальные затраты обеспечения прироста объема ее выпуска. Это объясняется тем, что выпуск других отраслей материалов и оборудования, увеличивающего производство этой отрасли, потребует в рамках межотраслевых связей дополнительных поставок с ее стороны.

Чтобы отразить эффект роста отрасли, в ее балансовом уравнении также присутствуют дополнительные поставки ее продукции с учетом прироста мощностей.

Распределение косвенных эффектов по годам реализации проекта осуществляется на базе информации, предоставляемой ФЭМ. Для перевода значений косвенных эффектов в форму,

сопоставимую с показателями ФЭМ, применяют показатели выпуска продукции проекта в каждый момент времени.

Распределение величины косвенных эффектов по годам здесь – это отношения объемов выпуска в регионе в итоге реализации проекта в конкретном году и выпуска продукции в прошлые периоды времени, определяемые в ФЭМ. Также используются косвенные эффекты, возникающие за счет реализации проекта в последнем году анализируемого периода и в предшествующие периоды времени, определяемые в ОММ.

Тогда указанный эффект можно использовать для поиска экономической эффективности проекта. Денежный поток для оценки эффективности проекта представлен здесь как переход от финансового анализа к экономическому в линейном балансирующем соотношении.

В нем определяется сальдо потоков денежных средств по инвестиционному проекту отдельного региона в период, предшествующий году реализации проекта. Оно применяется в анализе экономической (общественной) эффективности. Происходит суммирование аналогичного типа сальдо, но используемого для анализа финансовой (коммерческой) эффективности, дополнительных налоговых поступлений от проекта, субсидий на его осуществление. Эти денежные потоки применяют для расчета косвенных региональных и внешних эффектов проекта, а также ценовых эффектов, порожденных проектом.

В качестве региональной модели МОБ в [11] представлена региональная межотраслевая модель Республики Саха (Якутия), основанная на модели «затраты – выпуск». Модель содержит ряд видов экономической деятельности и соответствующих продуктов и описывает регион, состоящий из определенного набора муниципальных образований.

Соотношения финансовой части региональной межотраслевой модели строились с учетом неоднородности экономических процессов внутри регионов и связей с остальной страной и миром. Указанное обусловило следующую двухэтапную схему формализации данных процессов, осуществляемую последовательно.

1-ый этап. Распределение валовой добавленной стоимости (ВДС) по отраслям с учетом внутрирегиональной структуры имеет формальное описание в виде балансового уравнения. В нем ВДС, созданная и полученная производителями в регионе для отдельной отрасли, есть разность выпуска ее продукции в регионе и промежуточного потребления этой продукции другими отраслями.

При этом дополнительно к ВДС в регионе по конкретному виду продукции может также определяться как сумма ВДС по районам моделируемого региона. Они же, в свою очередь, получены произведением доли ВДС по продукции данного вида в этом районе на объем самого выпуска в регионе.

2-ой этап. Создание ВДС по элементам в регионе, предусматривающее агрегирование следующих региональных показателей. Это совокупность: заработной платы занятых в отрасли, валовая прибыль производителей и амортизация выпуска отрасли, объем налогов, уплаченных в бюджет региона всеми его субъектами при производстве, реализации и приобретении продукции отрасли, а также объем налогов, уплаченных в федеральный бюджет всеми субъектами региона при производстве, реализации и приобретении продукции отрасли.

3. Постановка задачи

В блоке межотраслевого финансового баланса модели АТМО предложена методология, содержащая ряд улучшений по сравнению с подходами, предлагаемыми типовыми моделями МОБ для финансовых потоков, сопутствующих деятельности *секторов*.

Указанный термин применен в работе (кроме обзора) вместо термина «*отрасль*» в силу его большей универсальности. Существующие подходы состоят, по итогам обзора, сделанного выше, в следующем:

– учет затрат только на продукцию, произведенную другими отраслями, но не ввезенную из-за пределов региона;

– недостаточное внимание, уделяемое всей совокупности налоговых платежей, с анализом экономического эффекта от предоставления налоговых преференций;

– почти полное отсутствие такого инструмента государственного регулирования, как субсидии и госфинансирование, предоставляемые отдельным секторам;

– принятие в расчет наличия кратковременных и долговременных кредитных выплат при анализе финансовых потоков секторов;

– недостаточный учет межрегиональных и внешнеторговых операций;

– почти полное игнорирование моделирования финансовых потоков для цепочек поставок.

Построенный блок межотраслевого финансового баланса направлен на устранение упомянутых выше пробелов. Это позволит четко очертить рамочные условия работы производственных и торговых секторов в смысле выполнения ими своих финансовых обязательств.

Решаемая в работе задача состоит в следующем. Объектом исследования является система m регионов с n секторами. Они соответствуют 456 видам экономической деятельности из ОКВЭД, содержанием которых является производство, с включением торговых предприятий.

Секторы согласно модели продукционного межотраслевого баланса региона [1] взаимодействуют в рамках выпуска продукции и ее ввоза/вывоза.

Внешнеэкономическая деятельность регионов описана в других блоках модели АТМО. В данной работе ставится цель отразить финансовые потоки, возникающие в результате межсекторного взаимодействия, наличия собственных и заемных денежных средств, а также выплат государственным и негосударственным организациям – для одного региона r.

Тогда на основе полученных из блока продукционного баланса [1] равновесных объемов:

- выпуска $x_1, x_2, \ldots, x_n$;
- ввоза $x^{(v)}_1, x^{(v)}_2, \ldots, x^{(v)}_n$;

(для простоты опускается индекс региона r), а также ряда других известных параметров (рис. 2) нужно определить наборы равновесных значений:

- отпускных цен выпуска $p_1, p_2, \ldots, p_n$;
- закупочных цен ввоза $p^{(v)}_1, p^{(v)}_2, \ldots, p^{(v)}_n$;

для которых удовлетворяются требования по операционным прибылям всех секторов региона $u^{(f)}_1, u^{(f)}_2, \ldots, u^{(f)}_n$ и ряд рамочных условий развития экономики региона, которые будут приведены ниже.

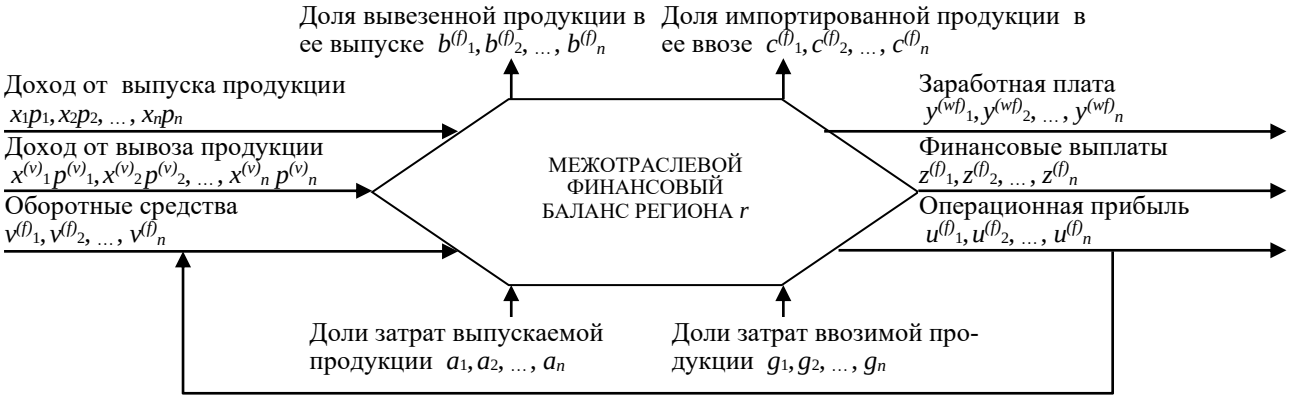


Рис. 2. Общая схема взаимодействия финансовых потоков в регионе r

Эта диаграмма разделена далее на ряд секторов с отражением разветвления по ним денежных потоков с учетом долей межсекторного использования их товаров в выпуске и ввозе (рис. 3).

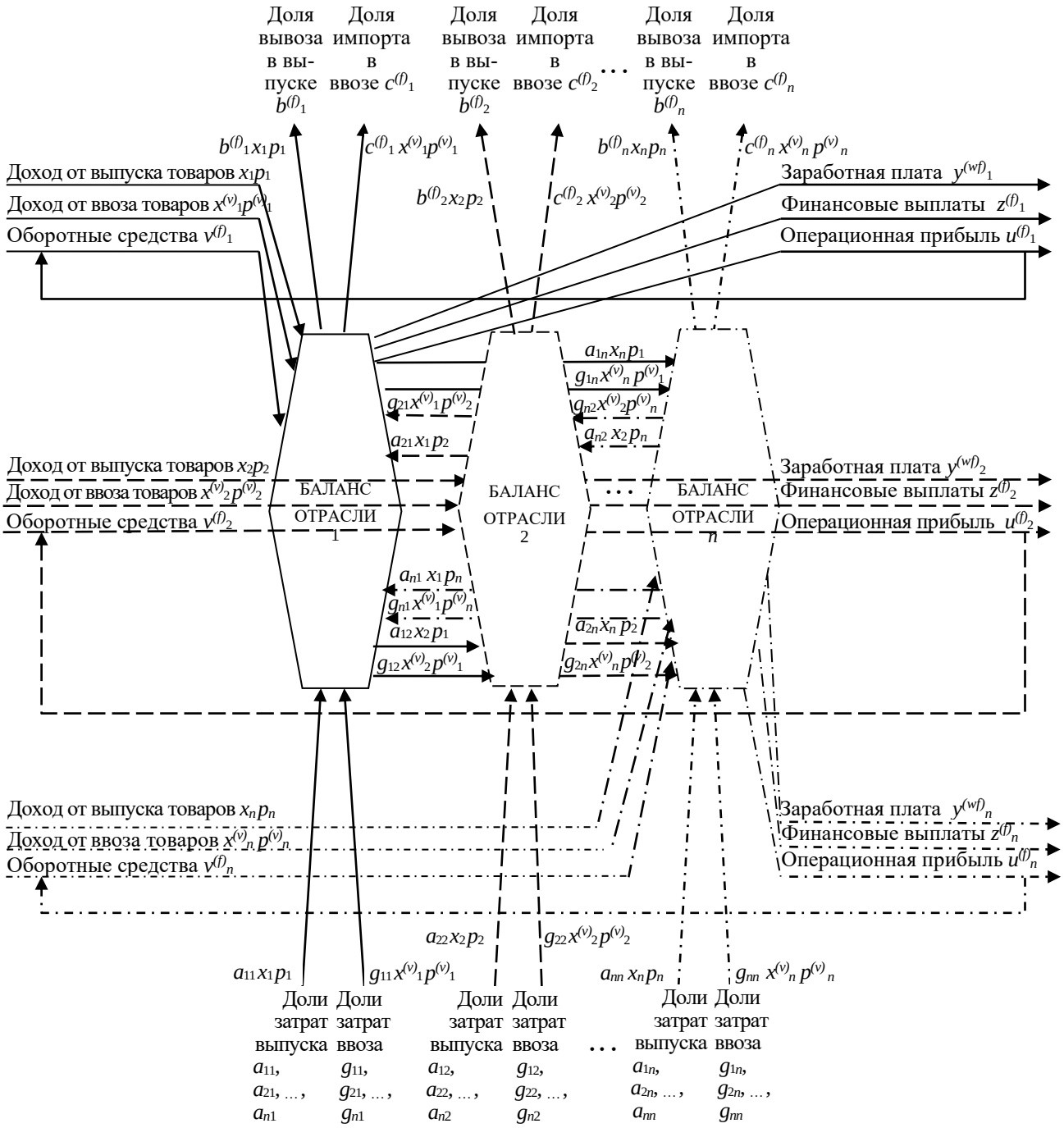


Рис. 3. Схема распределения денежных потоков в экономике региона

Схема движения финансовых потоков, откуда получается решение указанной задачи – это баланс каждого сектора между следующими компонентами. С одной стороны, это выручка от продукции – собственной и завезенной в регион, с другой – заработная плата и финансовые выплаты. Также на эту ситуацию влияют промежуточные затраты на выпуск и ввоз продукции других секторов и исключение той части продукции, которая реализуется за пределами региона по более высоким ценам. Переход параметров дохода на входе в целевые параметры денежных потоков на выходе модели определен технологическими и логистическими долями потребления секторами выпущенных и ввезенных товаров других секторов.

4. Строение финансового баланса модели АТМО

Описываемый в данной работе межсекторный финансовый баланс региона *r* для *n* секторов имеет следующий матричный вид:

$$ {}^{t}V^{(p)r} + {}^{t}X^{(p)r\backprime}\,{}^{t}A^{(f)r} + {}^{t}X^{(vpv)r\backprime}\,{}^{t}G^{(f)r} = {}^{t}Y^{(wf)r} + {}^{t}Z^{(f)r} + {}^{t}U^{(f)r}, \quad (1) $$

или, в натуральном выражении:

$$ {}^{t}V^{(f)r} + {}^{t}X^{r\backprime}\,{}^{t}P^{r}\,{}^{t}A^{(f)r} + {}^{t}X^{(v)r\backprime}\,{}^{t}P^{(v)r}\,{}^{t}G^{(f)r} = {}^{t}Y^{(wf)r} + {}^{t}Z^{(f)r} + {}^{t}U^{(f)r}, \quad (2) $$

где

- ${}^{t}X^{r\backprime}$ – диагональная матрица выпуска продукции в регионе;
- ${}^{t}X^{(v)r\backprime}$ – диагональная матрица ввоза продукции в регионе;
- ${}^{t}P^{r}$ – отпускные цены на выпускаемую продукцию;
- ${}^{t}P^{(v)r}$ – закупочные цены на ввозимую продукцию;
- ${}^{t}V^{(f)r}$ – оборотные средства;
- ${}^{t}Y^{(wf)r}$ – фонд зарплаты;
- ${}^{t}Z^{(f)r}$ – финансовые выплаты;
- ${}^{t}U^{(f)r}$ – операционная прибыль;
- ${}^{t}A^{(f)r}$ – технологические затраты;
- ${}^{t}G^{(f)r}$ – логистические затраты.

Основу параметров ${}^{t}A^{(f)r}$ и ${}^{t}G^{(f)r}$ составляют коэффициенты матриц ${}^{t}A^{r\mathrm{T}}$ и ${}^{t}G^{r\mathrm{T}}$. Они состоят из материальных издержек i-го сектора на единицу выпускаемой и ввозимой продукции j-го сектора ${}^{t}a_{ji}{}^{r}$ и ${}^{t}g_{ji}{}^{r}$. Сумма их произведения на объемы выпуска и ввоза ${}^{t}x^{(p)}_{j}{}^{r}$ и ${}^{t}x^{(vp)}_{j}{}^{r}$ в итоге дает затраты на выпуск и ввоз продукции i-го сектора.

Развернутый вид блока из n уравнений имеет следующий матричный вид:

$$ \underbrace{{}^{t}V^{(f)r} + {}^{t}X^{r}\,{}^{t}P^{r\backprime}}\,\underbrace{[(I - {}^{t}B^{(f)r}) - {}^{t}A^{r\mathrm{T}}]}_{{}^{t}A^{(f)r}} + \underbrace{{}^{t}X^{(v)r}\,{}^{t}P^{(v)r\backprime}}\,\underbrace{[(I - {}^{t}C^{(f)r}) - {}^{t}G^{r\mathrm{T}}]}_{{}^{t}G^{(f)r}} = $$

$$ = {}^{t}Y^{(wf)r} + {}^{t}Z^{(f)r} + {}^{t}U^{(f)r}. \quad (3) $$

- ${}^{t}B^{(f)}_{i}{}^{r}$ – доля выпускаемой продукции, реализованной вне региона;
- ${}^{t}C^{(f)}_{i}{}^{r}$ – доля ввозимой продукции, реализованной вне региона;
- ${}^{t}Y^{(wf)}_{i}{}^{r} = {}^{t\text{-}1}W^{(f)}_{i}{}^{r} \cdot {}^{t\text{-}1}Y^{(f)}_{i}{}^{r}$ – зарплата, определяемая числом работников и их средними зарплатами за (t-1)-ый месяц (зарплату выплачивают в первый день следующего месяца).

Тогда уравнение блока для сектора i, региона r и месяца t имеет вид:

$$ {}^{t}v^{(f)}_{i}{}^{r} + (1 - {}^{t}b^{(f)}_{i}{}^{r}) \cdot {}^{t}x_{i}^{r} \cdot {}^{t}p_{i}^{r} - {}^{t}x_{i}^{r} \cdot {}^{t}p_{i}^{r} \cdot \sum_{j=1}^{n} {}^{t}a_{ji}{}^{r} + (1 - {}^{t}c^{(f)}_{i}{}^{r}) \cdot {}^{t}x^{(v)}_{i}{}^{r} \cdot {}^{t}p^{(v)}_{i}{}^{r} - $$

$$ - {}^{t}x^{(v)}_{i}{}^{r} \cdot {}^{t}p^{(v)}_{i}{}^{r} \cdot \sum_{j=1}^{n} {}^{t}g_{ji}{}^{r} = {}^{t}y^{(wf)}_{i}{}^{r} + {}^{t}z^{(f)}_{i}{}^{r} + {}^{t}u^{(f)}_{i}{}^{r}. \quad (4) $$

Компоненты матрицы финансовых выплат ${}^{t}Z^{(f)r}$, используемые в выражениях как ${}^{t}z^{(f)}_{i}{}^{r}$, формируются из компонент выплат по кратко- и долгосрочным кредитам, ${}^{t}z^{(crd)}_{i}{}^{r}$ и ${}^{t}z^{(inv)}_{i}{}^{r}$:

$$ {}^{t}z^{(crd)}_{i}{}^{r} = {}^{t}v^{(crd)}_{i}{}^{r} \cdot \frac{{}^{t}b^{(x)}_{i}{}^{r}/12}{1 - \frac{1}{(1 + {}^{t}b^{(x)}_{i}{}^{r}/12)^{12}}} \cdot 12, \quad {}^{t}z^{(inv)}_{i}{}^{r} = {}^{t}v^{(inv)}_{i}{}^{r} \cdot \frac{{}^{t}b^{(z)}_{i}{}^{r}/12}{1 - \frac{1}{(1 + {}^{t}b^{(z)}_{i}{}^{r}/12)^{d/12}}} \cdot 12, \quad (5) $$

– здесь имеют место равномерные платежи ${}^{t}z^{(crd)}_{i}{}^{r}$ по краткосрочному кредиту ${}^{t}v^{(crd)}_{i}{}^{r}$ выплачиваемые каждый месяц в течение года, как и равномерные платежи ${}^{t}z^{(inv)}_{i}{}^{r}$ по долговременному кредиту ${}^{t}v^{(inv)}_{i}{}^{r}$. Период, на который он выдан (d), – един для всех участников рынка.

Добавим ряд следующих соотношений, где встречаются показатели прошлого месяца (t-1):

$$ {}^{t}b^{(f)}_{i}{}^{r} = (1 + {}^{t}db^{(f)}_{i}{}^{r}) \cdot \frac{{}^{t-1}z_{i}{}^{r}}{{}^{t-1}x_{i}{}^{r}}, \quad (6) $$

• доля вывоза продукции ${}^{t\text{-}1}z_{i}{}^{r}$ в ее выпуске ${}^{t\text{-}1}x_{i}{}^{r}$ с учетом роста экспорта продукции ${}^{t}db^{(f)}_{i}{}^{r}$;

$$ \left.\begin{array}{l} {}^{t}dv^{(f)}_{i}{}^{r} = {}^{t}v^{(crd)}_{i}{}^{r} + {}^{t}v^{(inv)}_{i}{}^{r} + {}^{t}v^{(mon)}_{i}{}^{r} + {}^{t}v^{(sub)}_{i}{}^{r} + {}^{t}v^{(fin)}_{i}{}^{r} \\ ({}^{t}v^{(inv)}_{i}{}^{r} = 0, \quad {}^{t}v^{(fin)}_{i}{}^{r} = 0, \quad t\,\%12 \neq 0), \quad (\% \ - \text{ остаток от деления}) \end{array}\right\} \quad (7) $$

• оборотные средства: краткосрочный кредит (ежемесячно) ${}^{t}v^{(crd)}{}_{i}^{r}$, долгосрочный кредит (ежегодно) ${}^{t}v^{(inv)}{}_{i}^{r}$, денежные средства ${}^{t}v^{(mon)}{}_{i}^{r}$, субвенции ${}^{t}v^{(sub)}{}_{i}^{r}$, госфинансирование ${}^{t}v^{(fin)}{}_{i}^{r}$;

$$ {}^{t}v^{(mon)}{}_{i}^{r} = (1 - {}^{t}d^{(pay)}{}_{i}^{r}) \cdot {}^{t-1}u^{(fre)}{}_{i}^{r}, \tag{8} $$

• денежные средства в виде доли от прибыли ${}^{t-1}u^{(fre)}{}_{i}^{r}$ после отчислений корпорации ${}^{t}d^{(pay)}{}_{i}^{r}$;

$$ {}^{t}p_{i}^{r} = (1 + {}^{t}dp_{i}^{r}) \cdot (1 + {}^{t}d^{(u)}{}_{i}^{r} + {}^{t}d^{(av)}{}_{i}^{r}) \cdot {}^{t}p^{(0)}{}_{i}^{r}, \tag{9} $$

• региональные цены выпуска на основе отпускных цен с учетом инфляции, доходности и НДС;

Параметры ${}^{t}b^{(f)}{}_{i}^{r}$ и ${}^{t}c^{(f)}{}_{i}^{r}$ извлекают из выручки часть, реализуемую вне региона – вывоз и импорт. Рост этих параметров определен следующим:

• ${}^{t}db^{(f)}{}_{i}^{r}$ – изменение доли экспорта за счет изменения вывоза ${}^{t}z_{i}^{r}$, его содержащего, равного ${}^{t}q_{i}^{rr} \cdot {}^{t}z_{i}^{r}$, где ${}^{t}q_{i}^{rr}$ – диагональный элемент матрицы перевозок вывоза торгового баланса, т.е. объем вывоза ${}^{t}z_{i}^{r}$ изменяется на величину ${}^{t}db^{(f)}{}_{i}^{r} \cdot {}^{t}q_{i}^{rr} \cdot {}^{t-1}z_{i}^{r}$;

• ${}^{t}dc^{(f)}{}_{i}^{r}$ – изменение доли прямого импорта за счет изменения импорта ${}^{t}z^{(v)}{}_{i}^{r}$, его содержащего, равного ${}^{t}o_{i}^{rr} \cdot {}^{t}z^{(v)}{}_{i}^{r}$ где ${}^{t}o_{i}^{rr}$ – диагональный элемент матрицы транзита импорта торгового баланса, т.е. импорт ${}^{t}z^{(v)}{}_{i}^{r}$ изменяется на величину ${}^{t}dc^{(f)}{}_{i}^{r} \cdot {}^{t}o_{i}^{rr} \cdot {}^{t-1}z^{(v)}{}_{i}^{r}$.

Уменьшение затрат ${}^{t}dg_{ji}^{r}$ на ввозимую продукцию ${}^{t}x^{(v)}{}_{i}^{r}$ других секторов достигается за счет доходов от ее продажи (по закупочным ценам ${}^{t}p^{(v)}{}_{i}^{r}$):

$$ {}^{t}x^{(v)}{}_{i}^{r} \cdot {}^{t}p^{(v)}{}_{i}^{r} \cdot \sum_{j=1}^{n} {}^{t}dg_{ji}^{r} \le {}^{t}v^{(d)}{}_{i}^{r} \cdot {}^{t}x^{(v)}{}_{i}^{r} \cdot {}^{t}p^{(v)}{}_{i}^{r}, \tag{10} $$

откуда

$$ \sum_{j=1}^{n} {}^{t}dg_{ji}^{r} \le {}^{t}v^{(d)}{}_{i}^{r}, \tag{11} $$

где ${}^{t}v^{(d)}{}_{i}^{r}$ – темп роста подушевого ВРП в качестве коэффициента воспроизводства.

Фактические затраты на логистику ${}^{t}p^{(g)}{}_{i}^{r}$ распределены пропорционально объемам продукции сектора j. Они используются логистическими службами для обеспечения ввоза единицы продукции сектора i, то есть пропорционально ранее введенным в [1] логистическим коэффициентам ${}^{t}g_{ji}^{r}$. При этом указанные затраты умножены на эффективность их применения: ${}^{t}c^{(ef)}{}_{i}^{r}$. Чем выше этот коэффициент, тем больше продукции других отраслей будут задействованы в логистических операциях сектора i.

Фактические затраты на логистику приводят к оптимизации работы торговых предприятий секторов (компонента $({}^{t}p^{(v)}{}_{i}^{r})^{-1}$ нужна для отражения доли ${}^{t}p^{(g)}{}_{i}^{r}$ в отпускной цене):

$$ {}^{t}dg_{ji}^{r} = {}^{t}p^{(g)}{}_{i}^{r} \cdot ({}^{t}p^{(v)}{}_{i}^{r})^{-1} \cdot {}^{t}g_{ji}^{r} \cdot {}^{t}c^{(ef)}{}_{i}^{r}. \tag{12} $$

Подставим это в условия (11) и выразим фактические удельные затраты на логистику:

$$ {}^{t}p^{(g)}{}_{i}^{r} \le {}^{t}v^{(d)}{}_{i}^{r} \cdot {}^{t}p^{(v)}{}_{i}^{r} \cdot \sum_{j=1}^{n} ({}^{t}dg_{ji}^{r} \cdot {}^{t}c^{(ef)}{}_{i}^{r})^{-1}. \tag{13} $$

Выразим затраты на улучшение логистики ${}^{t}p^{(g)}{}_{i}^{r}$ через прирост правой части неравенства, ограничивающего этот показатель, на величину ${}^{t}dp^{(g)}{}_{i}^{r}$:

$$ {}^{t}p^{(g)}{}_{i}^{r} = (1 + {}^{t}dp^{(g)}{}_{i}^{r}) \cdot {}^{t}v^{(d)}{}_{i}^{r} \cdot {}^{t}p^{(v)}{}_{i}^{r} \cdot \sum_{j=1}^{n} ({}^{t}dg_{ji}^{r} \cdot {}^{t}c^{(ef)}{}_{i}^{r})^{-1}. \tag{14} $$

Здесь рост ${}^{t}dp^{(g)}{}_{i}^{r}$ обеспечен базовыми удельными затратами ${}^{t}p^{(gt)}{}_{i}^{r}$ на логистику, косвенно ограничивающими фактические затраты ${}^{t}p^{(g)}{}_{i}^{r}$, что дает:

$$ {}^{t}dp^{(g)}{}_{i}^{r} \le {}^{t}p^{(gt)}{}_{i}^{r}. \tag{15} $$

Прирост капитала ${}^{t}v^{(f)}{}_{i}^{r}$ отражает сумму расходов по коэффициентам ${}^{t}da_{ji}^{r}$ и ${}^{t}dg_{ji}^{r}$ на производственные блага (для секторов, которые поставляют оборудование) и дополнительное сырье:

$$ {}^{t}v^{(f)}{}_{i}^{r} = {}^{t}x_{i}^{r} \cdot {}^{t}p_{i}^{r} \cdot \sum_{j=1}^{n} {}^{t}da_{ji}^{r} + {}^{t}x^{(v)}{}_{i}^{r} \cdot {}^{t}p^{(v)}{}_{i}^{r} \cdot \sum_{j=1}^{n} {}^{t}dg_{ji}^{r} \cdot \quad (16) $$

Указанный показатель отражает конкретное наполнение инвестиционных затрат в виде сырья и оборудования. Здесь источником денежных потоков являются:

• оборотные средства ${}^{t}v^{(f)}{}_{i}^{r}$;

• выручка от сбыта выпускаемой продукции ${}^{t}x_{i}^{r} \cdot {}^{t}p_{i}^{r}$;

• выручка от сбыта ввозимой продукции ${}^{t}x^{(v)}{}_{i}^{r} \cdot {}^{t}p^{(v)}{}_{i}^{r}$.

Оборотные средства ${}^{t}v^{(f)}{}_{i}^{r}$ формируются из денежных средств ${}^{t}v^{(mon)}{}_{i}^{r}$, краткосрочного ${}^{t}v^{(crd)}{}_{i}^{r}$, долгосрочного ${}^{t}v^{(inv)}{}_{i}^{r}$ кредитов. Первый обеспечивает текущее месячное финансирование, а второй задается как источник инвестиций, осуществляемых раз в год.

Долговременный кредит рассматривается в модели АТМО как основное средство реализации долговременных планов конкретного предприятия сектора. Эти планы заключаются в покупке нового промышленного и торгового оборудования, чья эффективность влияет на размер технологических и логистических коэффициентов соответственно. Для простоты принимается, что денежные средства долговременного кредита тратятся в первый месяц его получения, в отличие от ежемесячных кратковременных кредитов, покрывающих часть затрат предприятия. Они задаются как заемная часть оборотных средств в начале планирования и погашаются ежемесячными платежами.

Потребность в долгосрочном и краткосрочном кредитах может частично компенсироваться соответственно *субвенциями* и *госфинансированием* (раз в год), ${}^{t}v^{(sub)}{}_{i}^{r}$ и ${}^{t}v^{(fin)}{}_{i}^{r}$ – тогда размеры указанных кредитов снижаются.

Выпускаемая и ввозимая продукция секторов в объемах ${}^{t}x_{i}^{r} \cdot {}^{t}p_{i}^{r}$ и ${}^{t}x^{(v)}{}_{i}^{r} \cdot {}^{t}p^{(v)}{}_{i}^{r}$ сбывается по отпускным и закупочным ценам ${}^{t}p_{i}^{r}$ и ${}^{t}p^{(v)}{}_{i}^{r}$. Это себестоимости выпуска сектора ${}^{t}p^{(0)}{}_{i}^{r}$ с добавленной стоимостью ${}^{t}d^{(av)}{}_{i}^{r}$ и нормой доходности ${}^{t}d^{(u)}{}_{i}^{r}$. Ввезенные товары реализуются по закупочным ценам ввоза ${}^{t}p^{(v)}{}_{i}^{r}$ (средние цены рынка с торговой наценкой ${}^{t}dp^{(v)}{}_{i}^{r}$).

Далее из этой выручки вычитаются расходы предприятий сектора на продукцию других секторов региона ${}^{t}x_{i}^{r} \cdot {}^{t}a_{ji}^{r} \cdot {}^{t}p_{i}^{r}$, а также ввоз продукции из других регионов ${}^{t}x^{(v)}{}_{i}^{r} \cdot {}^{t}g_{ji}^{r} \cdot {}^{t}p^{(v)}{}_{i}^{r}$.

Кроме того, вычитается доля вывоза и импорта соответственно при выпуске и ввозе: ${}^{t}z_{i}^{r} = {}^{t}x_{i}^{r} \cdot {}^{t}b^{(f)}{}_{i}^{r}$ и ${}^{t}z^{(v)}{}_{i}^{r} = {}^{t}x^{(v)}{}_{i}^{r} \cdot {}^{t}c^{(f)}{}_{i}^{r}$, т.е. объемы реализации вывоза и импорта вне региона.

Эти доходы от сбыта вывоза (в т.ч. экспорта) и транзитного импорта за вычетом цен ввозимых из других регионов товаров рассчитывают в блоке межрегионального торгового баланса, который также входит в модель АТМО.

Также есть расходы на фонд зарплаты ${}^{t}y^{(wf)}{}_{i}^{r}$, финансовых выплат ${}^{t}z^{(f)}{}_{i}^{r}$, формируемых из выплат краткосрочных кредитов с процентами ${}^{t}z^{(crd)}{}_{i}^{r}$ и долгосрочных кредитов с процентами ${}^{t}z^{(inv)}{}_{i}^{r}$ (дается раз в год), затрат на рекламу ${}^{t}z^{(rec)}{}_{i}^{r}$, выплат по ЕСН ${}^{t}z^{(ern)}{}_{i}^{r} \cdot {}^{t}c^{(soc)}{}_{i}^{r}$ с оплаты труда работников ${}^{t}z^{(ern)}{}_{i}^{r}$ в соответствии с ЕСН ${}^{t}c^{(soc)}{}_{i}^{r}$. Учитываются и платежи на прирост активов ${}^{t-1}dv^{(f)}{}_{i}^{r} \cdot {}^{t}b^{(cap)}{}_{i}^{r}$ с учетом налога на прирост капитала ${}^{t}b^{(cap)}{}_{i}^{r}$ и самого прироста капитала за прошлый месяц ${}^{t-1}dv^{(f)}{}_{i}^{r}$, затрат на обновление основного капитала. Они, в свою очередь, представляют собой фиксированные активы ${}^{t}v^{(fix)}{}_{i}^{r}$, умноженные на долю ежегодных выплат на восстановление основных фондов ${}^{t}b^{(cap)}{}_{i}^{r}$.

В итоге получается значение операционной прибыли ${}^{t}u^{(f)}{}_{i}^{r}$, при этом показателем динамики выступает прирост операционной прибыли ${}^{t}du^{(f)}{}_{i}^{r} = {}^{t}u^{(f)}{}_{i}^{r} - {}^{t-1}u^{(f)}{}_{i}^{r}$.

5. Общие ограничения на значения переменных финансового баланса

Финансовый блок модели АТМО предназначен для поиска оптимальных значений переменных отпускных цен выпуска ${}^t p_i^r$ и закупочных цен ввоза ${}^t p^{(v)}{}_i^r$, логистических коэффициентов ${}^t g_{ji}{}^r$, индикаторов плановых долей продукции: вывоза в выпуске ${}^t b^{(f)}{}_i^r$ и импорта в ввозе ${}^t c^{(f)}{}_i^r$.

Поиск данных показателей осуществляется совместно с поиском оптимальных значений переменных продукционного блока модели АТМО, описанного в [1]. На этот процесс влияют ограничения, определяемые денежными потоками, направляемыми секторам за выпускаемую и ввозимую ими продукцию, сегменту рабочей силы и государству за их услуги.

Указанные ограничения ориентированы на операционную прибыль ${}^t u^{(f)}{}_i^r$ как индикатор индустриализации:

$$ {}^t u^{(f)}{}_i^r = (1-{}^t b^{(f)}{}_i^r)\cdot {}^t x_i^r \cdot {}^t p_i^r - {}^t x_i^r \cdot {}^t p_i^r \cdot \sum_{j=1}^{n} {}^t a_{ji}{}^r + (1-{}^t c^{(f)}{}_i^r)\cdot {}^t x^{(v)}{}_i^r \cdot {}^t p^{(v)}{}_i^r - $$

$$ - {}^t x^{(v)}{}_i^r \cdot {}^t p^{(v)}{}_i^r \cdot \sum_{j=1}^{n} {}^t g_{ji}{}^r - {}^t y^{(wf)}{}_i^r - {}^t z^{(f)}{}_i^r + {}^t v^{(f)}{}_i^r = {}^t u^{(f)}{}_i^r ({}^t \mathbf{p}_i^r, {}^t \mathbf{p}^{(v)}{}_i^r). \quad (17) $$

Индикатор описывает задачу промышленного развития – рост операционной прибыли, часть которой направляется на увеличение выпуска:

$$ u^{(f)}{}_i^r ({}^t \mathbf{p}_i^r, {}^t \mathbf{p}^{(v)}{}_i^r) \to \max, \quad (18) $$

(на основе отпускных и закупочных цен).

1. Отпускные цены выпуска ограничены инфляцией и дефляцией их среднего уровня ${}^t d^{(in)}{}_i^r$:

$$ (1+{}^t d^{(in)}{}_i^r) \cdot {}^t p^{(\mathrm{mid})}{}_i^r \le {}^t p_i^r \le (1-{}^t d^{(in)}{}_i^r) \cdot {}^t p^{(\mathrm{mid})}{}_i^r, \quad (19) $$

где ${}^t p^{(\mathrm{mid})}{}_i^r$ – средние отпускные цены выпуска, определенные на основе тренда себестоимости сектора i за предшествующие годы.

2. Закупочные цены ввоза ограничены инфляцией/дефляцией их среднего уровня ${}^t d^{(in)}{}_i^r$:

$$ (1+{}^t d^{(in)}{}_i^r) \cdot {}^t p^{(\mathrm{vmid})}{}_i^r \le {}^t p^{(v)}{}_i^r \le (1-{}^t d^{(in)}{}_i^r) \cdot {}^t p^{(\mathrm{vmid})}{}_i^r, \quad (20) $$

где ${}^t p^{(\mathrm{vmid})}{}_i^r$ – средние закупочные цены ввоза за предшествующие годы (наценка ввоза в регион).

3. Прирост операционной прибыли не меньше безрисковой доходности ${}^t u^{(fd)}{}_i^r$ с учетом новых предприятий ${}^t dc^{(x)}{}_i^r$ (из продукционного баланса [1]):

$$ (1+{}^t dc^{(x)}{}_i^r) \cdot {}^t u^{(fd)}{}_i^r \cdot {}^{t-1} u^{(f)}{}_i^r \le {}^t du^{(f)}{}_i^r. \quad (21) $$

4. Прирост капитала обеспечен за счет роста денежных активов ${}^t v^{(fd)}{}_i^r$:

$$ {}^t dv^{(f)}{}_i^r \le (1+{}^t dc^{(x)}{}_i^r) \cdot {}^t v^{(fd)}{}_i^r \cdot {}^{t-1} v^{(f)}{}_i^r. \quad (22) $$

5. Фонд зарплаты (доходы работающих) не больше прироста капитала:

$$ {}^t dv^{(f)}{}_i^r \ge (1+{}^t dc^{(x)}{}_i^r) \cdot ({}^t w^{(fd)}{}_i^r \cdot {}^{t-1} w^{(f)}{}_i^r) \cdot ({}^t y^{(fd)}{}_i^r \cdot {}^{t-1} y^{(f)}{}_i^r), \quad (23) $$

- ${}^t w^{(fd)}{}_i^r$ – темпы роста численности среднего класса;
- ${}^t w^{(f)}{}_i^r$ – число занятых;
- ${}^t y^{(fd)}{}_i^r$ – темпы роста ВРП;
- ${}^{t-1} y^{(f)}{}_i^r$ – средняя зарплата за предшествующий месяц.

6. Уровень жизни работников поддерживает рост внешней торговли:

$$ {}^t db^{(f)}{}_i^r \cdot {}^t x_i^r \cdot {}^t p_i^r + {}^t dc^{(v)}{}_i^r \cdot {}^t x^{(v)}{}_i^r \cdot {}^t p^{(v)}{}_i^r \ge $$

$$ \ge (1+{}^t dc^{(x)}{}_i^r) \cdot ({}^t w^{(fd)}{}_i^r \cdot {}^{t-1} w^{(f)}{}_i^r) \cdot ({}^t y^{(f)}{}_i^r \cdot {}^{t-1} y^{(f)}{}_i^r), \quad (24) $$

- ${}^{t}db^{(f)}{}_{i}^{r}$ – рост доли экспорта в объеме выпуска;
- ${}^{t}dc^{(f)}{}_{i}^{r}$ – рост доли импорта в объеме ввоза.

6. Рамочные энергетические условия работы финансового баланса

Условия циркулирования финансовых потоков в регионе имеют энергетическую основу, подразумевающую поддержку успешными секторами ТЭК других, неэнергетических секторов, часто убыточных, но с большим технологическим и социально-экономическим весом.

С этой целью в модели АТМО *принципы связи ТЭК и экономики*, приведенные в работе [1], включают закономерности циркулирования финансовых потоков среди предприятий региона. Это осуществление предприятиями своей деятельности при участии государственного регулирования, что реализует следующие принципы (они нумеруются, начиная с п. 7, так как нумерация ограничения на значения переменных финансового баланса заканчиваются на п. 6 в предыдущем разделе статьи).

7. Принцип *роста удельных затрат на ввозное сырье и оборудование секторами ТЭК* для ввоза торговыми предприятиями сырья и оборудования должен сказываться на доле логистической экономии ${}^{t}dg_{ji}{}^{r}$. Эти переменные являются инструментами изменения элементов транспонированной в рамках финансового баланса матрицы логистических коэффициентов ${}^{t}g_{ji}{}^{r}$ (частью которых являются ${}^{t}dg_{ji}{}^{r}$). Поэтому их положительные значения для энергетического сектора i и ряда неэнергетических секторов j дают рост ввоза в регион торговыми сетями товаров этих секторов:

$$ {}^{t}dg_{ji}{}^{r} > 0, \quad i \in I^{(s)}, \quad j \notin I^{(s)}, \tag{25} $$

где

- $I^{(s)}$ = {64–71, 72–75, 76–78, 443–457} – множество индексов секторов ТЭК, введенных в [1] на основе номеров 456 видов экономической деятельности из ОКВЭД, выделенных для отражения в модели АТМО.

8. Принцип *снижения удельных расходов на ввозные энергоресурсы* выполняется с помощью доли логистической экономии ${}^{t}dg_{ji}{}^{r}$. Ее отрицательный прирост указывает на уменьшение объема продукции i-го энергетического сектора, потребленного j-ым сектором, не принадлежащим к топливно-энергетическому кластеру, что влияет на логистические коэффициенты ${}^{t}dg_{ji}{}^{r}$:

$$ {}^{t}dg_{ji}{}^{r} > 0, \quad i \notin I^{(s)}, \quad j \in I^{(s)}, \tag{26} $$

где

- $I^{(s)}$ = {64–71, 72–75, 76–78, 443–457} – множество секторов ТЭК (на основе ОКВЭД).

9. Принцип *аккумуляции финансовых средств в рамках секторов ТЭК* осуществляется за счет прироста капитала ${}^{t}dv^{(f)}{}_{i}^{r}$ для выпуска и ввоза продукции ТЭК. Для выпуска прирост капитала предприятий ТЭК отражает приросшие основные фонды после реализации необходимых для развития инвестиционных проектов. Для вывоза же прирост капитала в торговле энергоресурсами и энергией между регионами ничтожно мал вследствие сохранившейся с советских времен развитой инфраструктуры доставки энергии и энергоресурсов между регионами:

$$ {}^{t}dv^{(f)}{}_{i-n}{}^{r} > 0, \quad {}^{t}dv^{(f)}{}_{i+n}{}^{r} \approx 0, \quad i \in I^{(s)}, \tag{27} $$

где

- ${}^{t}dv^{(f)}{}_{i-n}{}^{r}$ – прирост капитала на производстве;
- ${}^{t}dv^{(f)}{}_{i+n}{}^{r}$ – прирост капитала в торговле.

10. Принцип *обеспечения прибыльности предприятиями ТЭК выше уровня рентных платежей* основан на операционной прибыли ${}^{t}du^{(f)}{}_{i}^{r}$ производственных компаний ТЭК (индекс i–n). Здесь предполагается, что ее рост при любых изменениях не должен быть перекрыт рентными выплатами:

$$\sum_{i \in I^{(s)}} {}^{t}du^{(f)}{}_{i-n}{}^{r} \le \sum_{i \in I^{(s)}} {}^{t}y^{(wc)}{}_{i}{}^{r}, \quad (28)$$

где

• ${}^{t}y^{(wc)}{}_{i}{}^{r}$ – рента как произведение объемов сверхприбыли от продажи природных ресурсов.

11. Принцип *неравномерности получения прибыли неэнергетическими секторами и секторами ТЭК* базируется на операционной прибыли ${}^{t}u^{(f)}{}_{i}{}^{r}$ как интереса производственных предприятий (индекс *i–n*). При этом сумма операционной прибыли как целевая функция растет не за счет секторов ТЭК, чья прибыль к прошлому периоду не выше прогноза роста их ресурсной базы:

$$\sum_{i \in I} {}^{t}du^{(f)}{}_{i-n}{}^{r} \to \max, \quad {}^{t}u^{(f)}{}_{i-n}{}^{r} \ge (1 + {}^{f}y^{(cd)}{}_{i}{}^{r}) \cdot {}^{t-1}u^{(f)}{}_{i}{}^{r}, \quad i \in I^{(s)}, \quad (29)$$

• ${}^{f}y^{(cd)}{}_{i}{}^{r}$ – темп роста исследований ресурсной базы за год *f*;

• I – множество всех секторов региона.

Ограничения 1–11 определяют характер развития экономики региона.

7. Отличительные характеристики представленной модели

Описанная модель имеет ряд различий в сравнении с подобными моделями, созданными ранее [2–11]:

• отражение реального сегмента экономики – секторов выпуска и торговли товарами;

• участие в сбыте как складов промышленных компаний (онлайн-продажи своих товаров), так и магазинов торговых сетей (для товаров из-за пределов региона, в т.ч. импорта);

• получение оборотных средств из прибыли/кредитов и субвенций/госфинансирования;

• учет в себестоимости выпуска материалов и комплектующих не только сектора базового региона, но и других регионов, а также полученных по импорту;

• исключение из формирования операционной прибыли доходов, получаемых от сбыта продукции сектором в другие регионы и за рубежом; так как ее стоимость выше, чем по региону, то окончательный учет этой разницы производится в другом блоке модели АТМО;

• разделение косвенных затрат и выплат по выпуску и ввозу продукции секторов региона, несущих его промышленными и торговыми предприятиями, на следующие категории:

– средняя зарплата с включенным подоходным налогом;

– ЕСН в составе пенсионных, социальных, федеральных и других взносов;

– затраты на выплаты кредитов, рекламу, прирост капитала и амортизацию;

• изменение логистических затрат за счет инвестиций в торговое оборудование;

• учет роста производства и торговли в ценах – отпускных выпуска и закупочных ввоза;

• включение в доходность и процентов по кредитам, и транспортного налога/акцизов;

• использование в уравнениях блока операционной прибыли в роли невязки между выручкой и затратами как в виде промежуточного потребления, так и финансовых выплат;

• поиск оптимальных значений переменных финансового баланса при макроэкономических ограничениях, отражающих главные социально-экономические тренды в регионах, а также рамочные условия влияния ТЭК на экономику региона.

8. Заключение

Указанные характерные черты модели дадут ее пользователю возможность анализировать изменения ряда параметров денежных потоков, возникающих в процессе производства и торговли секторов региона своей продукцией, путем решения следующих задач:

• отыскание таких величин отпускных и закупочных цен на выпускаемую и ввозимую продукцию производственных и сбытовых предприятий секторов, которые бы стимулирова-

ли рост промышленности и торговли в условиях конкуренции за финансовые ресурсы;

• получение параметров функционирования промышленных и торговых предприятий, способствующих большей эффективности использования ими финансовых средств, в т.ч. с помощью снижения логистических затрат ввоза продукции в регион;

• формирование для органов государственного управления стратегий стимулирования роста предприятий отраслей за счет снижения налогов, а также субвенций и госфинансирования;

• поиск параметров оптимальной структуры затрат предприятий торгового кластера, минимизирующих потребление товаров других секторов для ввоза в регион, размещения на складах/торговых точках и сбыта продукции, что ограничивает финансовые затраты секторов торговли, снижая цены на ввоз потребительских товаров;

• поддержка выработки обоснований для правительства и центрального банка по предоставлению ряду секторов конкретных регионов преференций по налоговым выплатам и снижению ставок банковских кредитов;

• выявление критических значений показателей финансовых потоков между секторами региона, при которых реализуемы условия, обеспечивающие рост экономики региона, в т.ч. промышленности, за счет ресурсов, предоставляемых секторами ТЭК, и не в ущерб последним.

Литература

1. *Самков Т. Л.* Продукционный баланс экономики региона // Вестник СибГУТИ. 2021. № 2 (54). С. 27–49.
2. *Татаркин Д. А., Сидорова Е. Н., Трынов А. В.* Использование матрицы финансовых потоков в моделировании влияния федеральной налогово-бюджетной политики на экономику региона (на примере Свердловской области) // Вестник УрФУ. Серия экономика и управление. 2014. № 6. С. 67–76.
3. *Демидько Е. В.* Круговые потоки как инструмент исследования социально-экономической системы региона // Фундаментальные исследования. 2015. № 6 (3). С. 566–575.
4. *Шелехова Л. В., Блягоз З. У., Нагоев А. В., Тешев В. А.* Межотраслевой баланс и модель «затраты – выпуск»: история создания и перспективы развития // Интернет-журнал «Науковедение». 2015. Т. 7, № 2, http://naukovedenie.ru/PDF/114EVN215.pdf.
5. *Машунин Ю. К., Машунин И. А.* Организация управления, моделирование и прогнозирование развития экономики региона // Региональная экономика и управление: электронный научный журнал, https://eee-region.ru/article/4503/.
6. *Турко В. А.* Модель оптимизации финансовых потоков в рамках промышленной кооперации стран ЕАЭС // Гипотеза. 2018. № 3 (4). С. 88–97.
7. Экономика Сибири в условиях глобальных вызовов XXI века: сборник статей в 6 томах. Том 4: Модели и методы исследований перспектив социально-экономического развития Сибири и России в целом и в разрезе отраслевых комплексов и макрорегионов / под ред. В. И. Суслова, Н. В. Горбачевой. Новосибирск: ИЭОПП СО РАН, 2018. С. 147–149.
8. *Потапенко В. В.* Оценка системы функций потребительского спроса вида «ПАДС» для России // Материалы конференции ИНП РАН и ИЭОПП СО РАН по межотраслевому и региональному анализу и прогнозированию «Экономическая политика России в межотраслевом и пространственном измерении», 21–22 марта 2019 г. Т. 1. С. 92–96.
9. *Куранов Г. О., Стрижкова Л. А., Тишина Л. И.* Межотраслевые и факторные модели в макроэкономическом анализе и межотраслевых исследованиях // Вопросы статистики. 2021. № 28 (2). С. 5–23.
10. Модели, анализ и прогнозирование пространственной экономики. Новосибирск: Изд-во ИЭОПП СО РАН, 2022. С. 169–178.
11. Модели и методы прогнозирования: Азиатская Россия в экономике страны. Новосибирск: Изд-во ИЭОПП СО РАН, 2023. С. 322–325.

Самков Тимур Леонидович

к.т.н., доцент кафедры математического моделирования и цифрового развития бизнес-систем Сибирского государственного университета телекоммуникаций и информатики (СибГУТИ, 630102, Новосибирск, ул. Кирова, д. 86), тел. +7 383 2698 278;

доцент кафедры производственного менеджмента и экономики энергетики Новосибирского государственного технического университета (НГТУ, 630073, Новосибирск, пр-т К. Маркса, д. 20), тел. +7 383 3461 359, e-mail: eerming@yandex.ru, ORCID ID: 0000-0001-6400-7672.

Автор прочитал и одобрил окончательный вариант рукописи.
Автор заявляет об отсутствии конфликта интересов.

Financial Balance of the Region's Economy

Timur L. Samkov

Siberian State University of Telecommunications and Information Science (SibSUTIS)
Novosibirsk State Technical University (NSTU)

Abstract: The article continues a series of papers describing the ATMO model (analysis of territorial multisectoral objects). It reflects the structure and functionality of the financial block of this model, which belongs to the Leontief input-output models. The block schematizes the mutual settlements of production and sales enterprises within the framework of mutual supplies of manufactured and imported products within the region. These monetary relations include accounting for the total cost of production and distribution, import and sale of products. They correspond to the production and sales companies of the region for all its sectors, followed by correlation with the revenue received from production and sales activities. In addition to accounting for the marketing component of the resulting operating profit of the sector, the novelty of the model consists in accounting for subsidies and state financing, reflecting the costs of products produced outside the region. Additionally, mechanisms have been introduced to account for credit and tax payments on the overall financial outcome of the sectors. Formal mechanisms have been developed for the relationship between investments in logistics and the costs of importing and selling non-regional goods and equipment. The model includes ratios that set the movement towards optimal values of those of its indicators, which, based on the principles of industrial development, provide for the availability of funds for the purchase of additional machines and equipment, increasing the volume of production and sales operations. These funds are available to the sector if its operating profit is greater than zero, and they are not available otherwise.

Keywords: input-output model, logistic coefficients, sector, segment, unit costs, cost, share of logistic overruns / savings.

For citation: Samkov T. L. Financial balance of the region's economy discourse (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 3, pp. 99-114. https://doi.org/10.55648/1998-6920-2024-18-3-99-114.

Content is available under the license Creative Commons Attribution 4.0 License

© Samkov T. L., 2024

The article was submitted: 06.04.2024;
revised version: 02.08.2024;
accepted for publication 09.08.2024.

References

1. Samkov T. L. Produkcionnyj balans ekonomiki regiona [Production balance region economy]. *Vestnik SibGUTI*, 2021, no. 2(54), pp. 27-49.
2. Tatarkin D. A., Sidorova E. N., Trynov A. V. Ispol'zovanie matricy finansovyh potokov v modelirovanii vlijanija federal'noy nalogovo-bjudzhetnoj politiki na ekonomiku regiona (na primere Sverdlovskoj oblasti) [The use of the financial flow matrix in modeling the impact of federal fiscal policy on the economy of the region (using the example of the Sverdlovsk region)]. *Vestnik* UrFU. Economics and Management Series, 2014, no. 6, pp. 67-76.
3. Demid'ko E.V. Krugovye potoki kak instrument issedovanija social'no-ekonomicheskoj sistemy regiona [Circular flows as a research tool for the socio-economic system of the region]. *Fundamental research*, 2015, no. 6(3), pp. 566-575.
4. Shelehova L. V., Bljagoz Z. U., Nagojev A. V., Teshev V. A. Mezhotraslevoy balans i model' "zatraty – vypusk": istorija sozdanija i perspektivy razvitija [Intersectoral balance and the input-output model: the history of creation and development prospects]. *Naukovedenie*, 2015, vol. 7, no. 2, available at: http://naukovedenie.ru/PDF/114EVN215.pdf.
5. Mashunin Ju. K., Mashunin I. A. Organizacija upravlenija, modelirovanie i prognozirovanie razvitija ekonomiki regiona [Management organization, modeling and forecasting of regional economic development]. *Regional economics and management: electronic scientific journal,* 2016, no.1(45), available at: https://eee-region.ru/article/4503/.
6. Turko V. A. Model' optimizacii finansovyh potokov v ramkah promyshlennoj kooperacii stran YeAES [A model for optimizing financial flows within the framework of industrial cooperation of the EAEU countries]. *Hypothesis*, 2018. no.3(4), pp. 88-97.
7. *Ekonomika Sibiri v uslovijah global'nyh vyzovov XXI veka: sbornik statej v 6-ti t. Tom 4: Modeli i metody issledovanij perspektiv social'no-ekonomicheskogo razvitija Sibiri i Rossii v celom i v razreze otraslevyh kompleksov i makroregionov* [The economy of Siberia in the context of global challenges of the XXI century: a collection of articles in 6 vol. Vol. 4: Models and methods of research on the prospects of socio-economic development of Siberia and Russia in general and in the context of industrial complexes and macro-regions]. Novosibirsk, IEOPP, Sibirskoe otdelenie Rossiiskoi akademii nauk, 2018, pp. 147-149.
8. Potapenko V. V. Ocenka sistema funkcij potrebitel'skogo sprosa vida «PADS» dlja Rossii [Evaluation of the system of consumer demand functions of the "PADS" type for Russia]. *Ekonomicheskaja politika Rossii v mezhotraslevom i prostranstvennom izmerenii: materialy konferencii INP Rossiiskoi akademii nauk IEOPP, Sibirskoe otdelenie Rossiiskoi akademii nauk po mezhotraslevomu i regional'nomu analizu i prognozirovaniju*, Moscow, 21-22 March, 2019, vol. 1, pp. 92-96.
9. Kuranov G. O., Strizhkova L. A., Tishina L. I. Mezhotraslevye i faktornye modeli v makroekonomicheskom analize i mezhotraslevyh issledovanijah [Intersectoral and factor models in macroeconomic analysis and intersectoral research]. *Voprosy statistiki*, 2021, no. 28(2), pp. 5-23.
10. *Modeli, analiz i prognozirovanie prostranstvennoj ekonomiki* [Models, analysis and forecasting of spatial economics], ed. V. I. Suslov, scientific ed. Ju. S. Jershov, Novosibirsk, IEOPP, Sibirskoe otdelenie Rossiiskoi akademii nauk, 2022, pp. 169-178.
11. *Modeli i metody prognozirovanija: Aziatskaja Rossija v ekonomike strany* [Forecasting models and methods: Asian Russia in the country's economy] ed. A. O. Baranov, V. I. Suslov, Novosibirsk, IEOPP, Sibirskoe otdelenie Rossiiskoi akademii nauk, 2023, pp. 322-325.

Timur L. Samkov

Cand. of Sci. (Engineering), Associate Professor of the Department of Mathematical Modeling and Digital Development of Business Systems, Siberian State University of Telecommunications and Information Science (SibSUTIS, Russia, 630102, Novosibirsk, Kirov St. 86), phone: +7 383 2698 278;

Associate Professor of the Department of Production Management and Energy Economics, Novosibirsk State Technical University (NSTU, Russia, 630073, Novosibirsk, av. K. Marksa, 20), phone: +7 383 3461 359, e-mail: eerming@yandex.ru, ORCID ID: 0000-0001-6400-7672.

DOI: 10.55648/1998-6920-2024-18-3-115-139 УДК 004.932.1:004.032.26

Исследование границ интенсивности видеопотока при FPV-управлении БПЛА в режиме предсказания кадров. Часть I: модели и методы *

А. А. Березкин, А. А. Ченский, Р. В. Киричек

Санкт-Петербургский гос. унив. телеком. им. проф. М. А. Бонч-Бруевича (СПбГУТ)

Аннотация: В настоящее время беспилотные летательные аппараты нашли широкое применение в различных сферах народного хозяйства. FPV-управление (управление от первого лица) относится к числу способов управления беспилотными летательными аппаратами, при котором видеопоток от беспилотного воздушного судна передается на станцию внешнего пилота в реальном масштабе времени. Из-за сбоев в работе сети связи пакеты с данными видеопотока могут потеряться или задержаться и быть доставленными с опозданием. Одним из способов, позволяющих компенсировать падение FPS в случае потери или задержки кадров на станции внешнего пилота, является прогнозирование промежуточных кадров видеопотока. В настоящей работе представлена схема предсказания промежуточных кадров видеопотока и приведены результаты экспериментов по установлению реально возможного увеличения FPS в нейросетевом кодеке, в котором для сжатия видео используются различные автокодировщики типа VQ-f16, а также алгоритм сжатия латентного пространства признаков без потерь DEFLATE, при этом для предсказания используется динамическая многомасштабная нейронная сеть воксельного потока DMVFN. Разработана регрессионная модель для прогнозирования времени предсказания. Формализована задача анализа FPS при различных конфигурациях нейросетевого декодера на стороне внешнего пилота.

Ключевые слова: FPS, вариационный автокодировщик, нейронный декодер, DMVFN, нейронные сети, предсказание видеопотока.

Для цитирования: Березкин А. А., Ченский А. А., Киричек Р. В. Исследование границ интенсивности видеопотока при FPV-управлении БПЛА в режиме предсказании кадров. Часть I: модели и методы // Вестник СибГУТИ. 2024. Т. 18, № 3. С. 115–139. https://doi.org/10.55648/1998-6920-2024-18-3-115-139.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Березкин А. А., Ченский А. А., Киричек Р. В. 2024

Статья поступила в редакцию 5.06.2024;
принята к публикации 20.06.2024.

1. Введение

В настоящее время беспилотные летательные аппараты (БПЛА) широко используются как в гражданской, так и в военной сферах [1]. При этом в случаях, когда полёт в автоматическом режиме невозможен или затруднителен, используется дистанционное управление БПЛА. Одной из разновидностей дистанционного управления БПЛА является управление от

* Научная статья подготовлена в рамках прикладных научных исследований СПбГУТ, регистрационный номер 1023031600087-9-2.2.4;2.2.5;2.2.6;1.2.1;2.2.3 в ЕГИСУ НИОКТР.

первого лица (first person view – FPV). FPV-управление предполагает передачу видеопотока и данных телеметрии от БПЛА на станцию внешнего пилота (СВП) по линии связи с землей (down link), а от СВП на БПЛА по линии связи с бортом (up link) – команд управления. Преимуществом такого способа управления является возможность управления БПЛА вне поля прямой видимости без вспомогательных наблюдательных средств, не установленных на БПЛА.

На текущий момент для осуществления связи между БПЛА и СВП, кроме радиолиний прямой видимости, широко используются мобильные сети связи. Тем не менее значительная часть территории Российской Федерации не покрыта мобильной связью. В перспективе обеспечить связь для управления БПЛА на всей территории Российской Федерации согласно Стратегии развития отрасли связи Российской Федерации на период до 2035 года [2] смогут гибридные орбитально-наземные сети связи (ГОНСС).

На конец 2023 года спутниковая группировка Российской Федерации насчитывала около 230 космических аппарата (КА) [3]. К 2036 году их количество планируется увеличить в 15 раз. При этом количество официально зарегистрированных на территории Российской Федерации беспилотных воздушных судов (БВС) на конец 2023 года уже исчислялось десятками тысяч (более 93 тысяч штук) и выросло с прошлого года на более чем 30 % [4]. Таким образом, существует значительная диспропорция между числом КА и БВС.

Можно предположить, что в перспективе 2035–2036 годов наличие задержек и потерь пакетов при FPV-управлении БПЛА прежде всего на отдалённых территориях Российской Федерации сохранится, т.к. только некоторая часть из 3435 КА, которые планируется вывести на орбиту к 2036 году, будет использоваться в качестве активных сетевых устройств ГОНСС, имеющей ограниченную пропускную способность.

Задержки и потери пакетов приводят к снижению частоты кадров в секунду (frame per second – FPS) видеопотока на стороне внешнего пилота (ВП), что может привести к ошибкам управления БПЛА от первого лица. В [5] предложен механизм адаптивного контроля интенсивности видеопотока при передаче FPV-трафика БВС на основе контрольных карт Шухарта и моделей временных рядов (рис. 1).

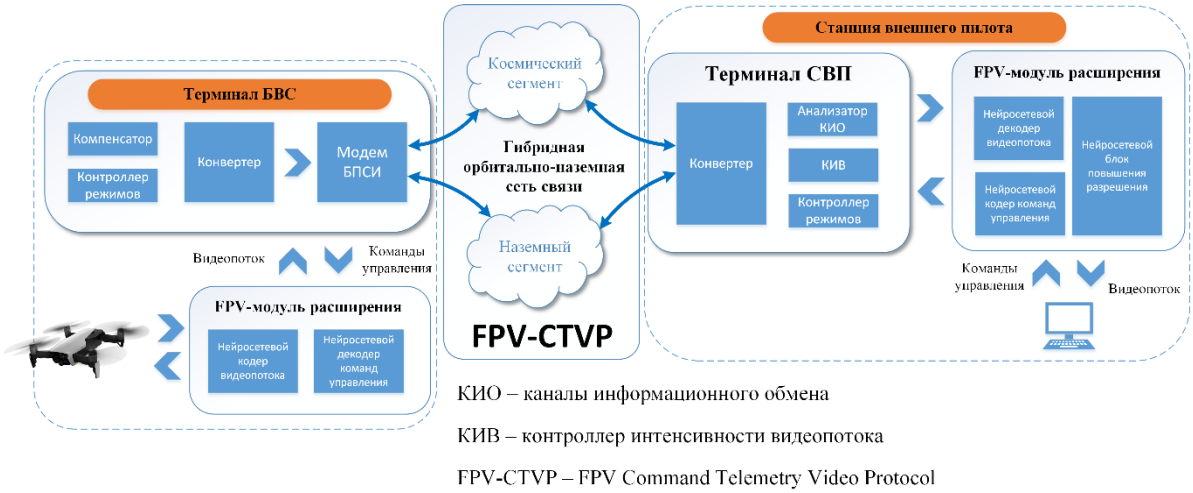


Рис. 1. Схема информационного обмена при FPV-управлении в ГОНСС

Система FPV-управления, представленная на рис.1, предназначена для предоставления услуги транспорта FPV-трафика с адаптивным контролем FPS видеопотока через различные сегменты инфраструктуры пакетных сетей связи и обеспечивает сокращение задержки передачи видеопотока и агрегацию команд управления с БВС на СВП.

На стороне СВП плавность отображения видеопотока оператору обеспечивается анализатором каналов информационного обмена (КИО), который контролирует реальный FPS на стороне оператора и отправляет команды в рамках специализированного протокола прикладного уровня FPV-CTVP (FPV-Command Telemetry Video Protocol), обеспечивающего пе-

редачу основных и служебных информационных потоков. В случае ухудшения качества канала (уменьшения ширины полосы пропускания) на стороне БВС обеспечивается выкалывание кадров видеопотока, которые не могут быть переданы вовремя, а на стороне СВП предсказание недостающих кадров модулем контроля интенсивности видеопотока (КИВ).

Ключевым элементом данной схемы является модуль КИВ, концепция и методы реализации которого рассмотрены в [6, 7], где предлагается использование нейросетевых моделей для предсказания последующих кадров видеопотока. Тем не менее проблема использования данных моделей в рамках модуля КИВ не является достаточно исследованной. Остаётся ряд вопросов, подлежащих изучению:

1. Время работы декодера t. Заданный уровень FPS определяется не только периодом поступления кадров на СВП, но и временем работы декодера. В случае использования для декодирования алгоритмов высокой сложности или нейросетевых моделей время исполнения становится слишком велико. В то же время в [6, 7] практически не рассматривается параметр времени выполнения моделей, а внимание уделяется метрикам качества восстанавливаемых кадров и уменьшению объёма видеопотока в целом.

2. Способность предсказательной модели обеспечивать увеличение FPS при некотором минимальном периоде поступления кадров $T_{\min}$. В случае ограниченности времени выполнения моделей всегда найдётся достаточно большой период $T_{\min}$, при котором модель способна обеспечить увеличение FPS.

3. Обоснованность использования различных моделей предсказания кадров. Целесообразной для использования признаётся такая модель, при которой минимальный период поступления кадров $T_{\min}$ меньше времени потери актуальности кадра $t_{\max}$. Данный параметр зависит от интенсивности смены окружения и различается для разных задач, решаемых БПЛА в рамках FPV-управления.

4. Использование механизмов интерполяции в потоке предсказания кадров. Это зависит от времени выполнения интерполяции и предсказания и от конкретных алгоритмов либо моделей машинного обучения, используемых для решения данных задач.

5. Предел возможного увеличения параметра кадров в секунду $\mathrm{FPS}_{\max}$. Так как выполнение нейросетевых моделей требует некоторого времени, FPS ограничено. Следовательно, возникает вопрос, насколько возможно максимально увеличить FPS с помощью некоторой предикторной модели.

6. Технические аспекты реализации модуля КИВ и его работы в рамках системы информационного обмена между БПЛА и СВП.

7. Определение параметра $t_{\max}$ для различных задач, возникающих в рамках FPV-управления БПЛА.

8. Выбор наилучшей предсказательной модели с точки зрения $\mathrm{FPS}_{\max}$. Определение компромиссного варианта модели с точки зрения времени выполнения и качества восстанавливаемых кадров в соответствии с требованиями обеспечения качества услуги.

В настоящей работе рассматриваются вопросы 1 – 6. Целью данной работы является исследование возможности практического применения предложенной концепции контроля интенсивности видеопотока [6, 7] в рамках нейросетевого декодера [8].

2. Экспериментальная установка

Для реализации экспериментальной установки используется прототип нейросетевого декодера (рис. 2).

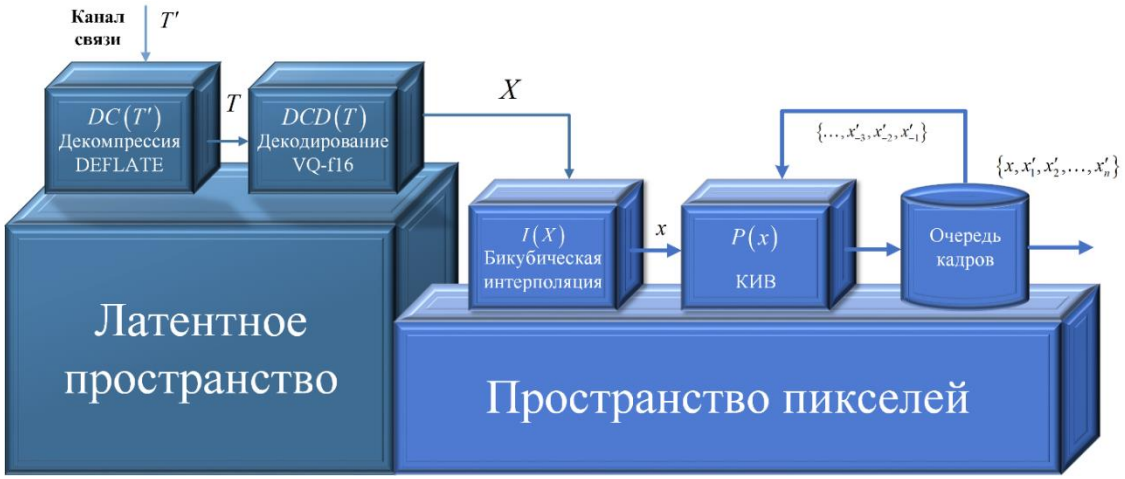


Рис. 2. Прототип нейросетевого декодера

Со стороны кодировщика, расположенного на борту БПЛА, передаются сжатые кадры видеопотока в латентном представлении T' в формате *fp16*, которые через канал связи между БПЛА и СВП поступают на блок декомпрессии *DC* нейросетевого декодера на стороне СВП. После декомпрессии кадры видеопотока в латентном пространстве T поступают на блок декодирования *VQ-f16* [9] *DCD*, на выходе которого получается пиксельное представление кадра X. Далее X поступает на блок бикубической интерполяции I для увеличения до некоторого разрешения x. На блок КИВ поступает как интерполированный кадр x, так и n ранее предсказанных кадров x_{-n}', …, x_{-3}', x_{-2}', x_{-1}'. На выход блока КИВ подаётся n+1 кадров: x и предсказанные x_1', x_2', x_3', …, x_n'. В качестве предсказательной модели блока КИВ используется нейросетевая модель DMVFN (Dynamic Multi-scale Voxel Flow Network) [10]. С выхода модуля КИВ кадры поступают в очередь и далее на оборудование ВП.

Модели нейросетевого декодера реализованы с помощью библиотеки PyTorch и выполняются на языке программирования Python версии 3.11. Для корректности все измерения времени выполнения отдельных блоков проводятся с использованием функций time_ns модуля стандартной библиотеки time и torch.cuda.syncronize(), что позволяет корректно измерять время выполнения каждого элемента блока, но может привести к небольшому увеличению общего времени работы ввиду асинхронности операций CUDA.

Входными параметрами нейросетевого декодера являются: разрешение поступающих кадров (a, x, b), число предсказываемых кадров n, результирующее разрешение (w, x, h).

Эксперименты в рамках настоящего исследования проводятся на серверной платформе Gigabyte G481-HA1. Модели нейронных сетей выполняются на видеокарте Nvidia Tesla A100 с 80 Гб видеопамяти.

3. Методика предсказания кадров видеопотока

Пусть некоторые кадры видеопотока x_i и x_{i+d} поступили в моменты времени i и (i+d). Методика предсказания кадров видеопотока заключается в нахождении кадров, которые бы поступили из канала связи в моменты времени между i и i+d (рис. 3, d = 4). Тем не менее такое определение не учитывает временные затраты на декомпрессию, декодирование и интерполяцию каждого кадра. Поэтому определим методику предсказания кадров видеопотока в контексте нейросетевого кодера другим образом.

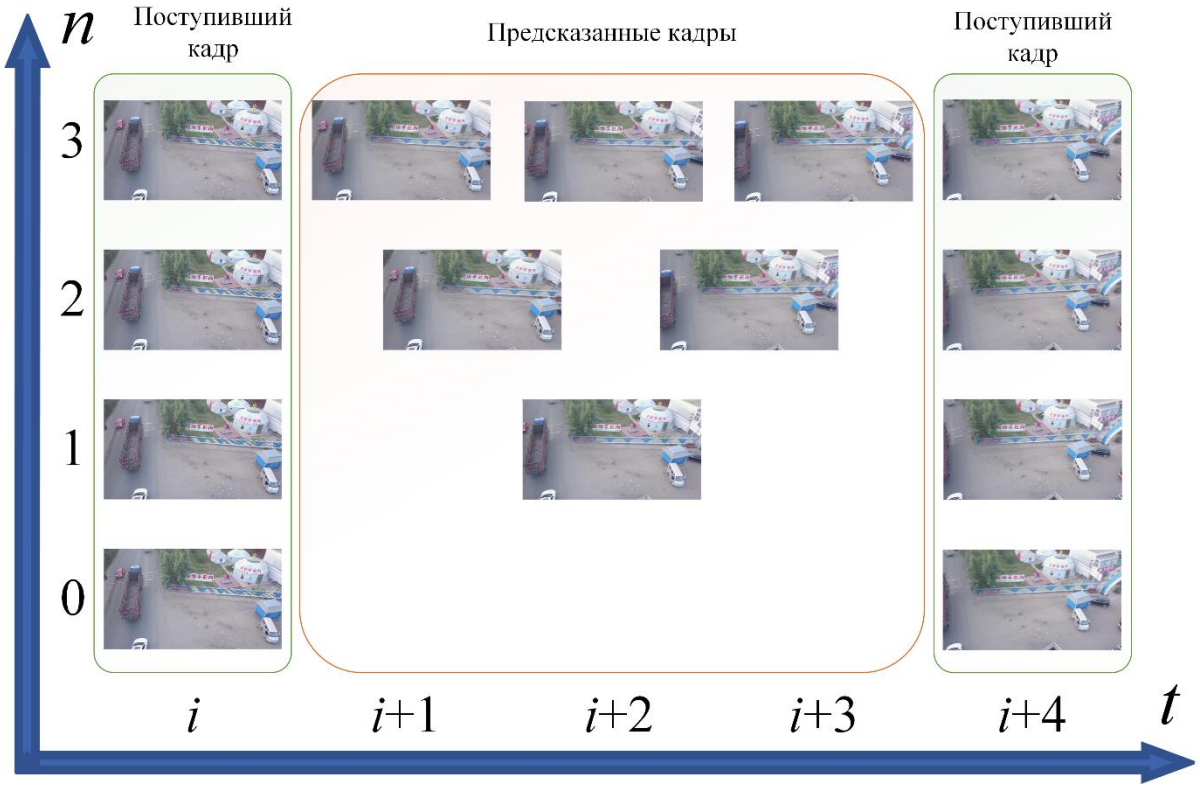


Рис. 3. Методика предсказания кадров видеопотока

Пусть z_i – суммарное время чтения из буфера, преобразования, декомпрессии, декодирования и интерполяции поступающего на нейросетевой декодер кадра, y_i – интервал времени между поступлением i и i+1 кадра, d_i – интервал времени между поступлением i-го кадра и некоторого следующего отображаемого кадра i' внешнему пилоту, t_i – время поступления i-го кадра. Поступающие кадры обрабатываются последовательно: $T_i', T_{i+1}', T_{i+2}', \ldots, T_{i+d}'$. Последний сжатый латентный кадр T_i', поступивший из канала связи, буферизируется и может быть прочитан сразу после завершения декодером работы над предыдущим. При каждом поступлении нового кадра T_{i+1}' буфер перезаписывается. Определим некоторый кадр в очереди кадров x_j* как потерявший актуальность в тот момент, когда в очередь кадров поступил новый кадр x_j. Соответственно, в один момент времени актуальный кадр x_j единственен. Далее определим J_t как множество индексов всех кадров, которые до некоторого времени t включительно были актуальными. Факт того, что все актуальные кадры поступают в очередь кадров, означает, что все они читались из буфера и обрабатывались декодером. Тогда актуальный кадр x_j в некоторый момент времени можно определить по формуле:

$$x_j : j : \left(\sum_{i=0}^{j} y_i = \left(\sum_{k \in J_t} z_k \right) - z_j \right), \tag{1}$$

а d_j – по формуле:

$$d_j = \begin{cases} 0, j \notin J_{tj} \\ \max\left(z_0, y_0\right), j \in J_{tj}, t < t_1 \\ z_j, j \in J_{tj}, t > t_1 \end{cases}.$$

Выражение (1) можно переписать в виде, где p – достаточно малая случайная величина:

$$x_j : j = \max(j) : \left(\sum_{i=0}^{j} y_i = \sum_{k \notin J_{tj+zj-p}} z_k \right), t_j \le t, p > 0, p \to 0,$$

$$x_j : j = \max(j) : \left(\sum_{i=0}^{j} y_i = t_j \right), t_j \le t .$$

Таким образом, методика предсказания кадров видеопотока заключается в предсказании n кадров на временном интервале (t_i+z_i; t_i+d_i+z_i'), таких что

$$t_{i:k} = t_i + z_i + \frac{d_i}{n} \times k,\ k = 1, 2, 3, ..., n ,$$

где $t_{i:k}$ – момент времени предсказания k-го кадра после i-го пришедшего с кодера кадра.

Процесс функционирования системы последовательной обработки кадров представлен на рис. 4.

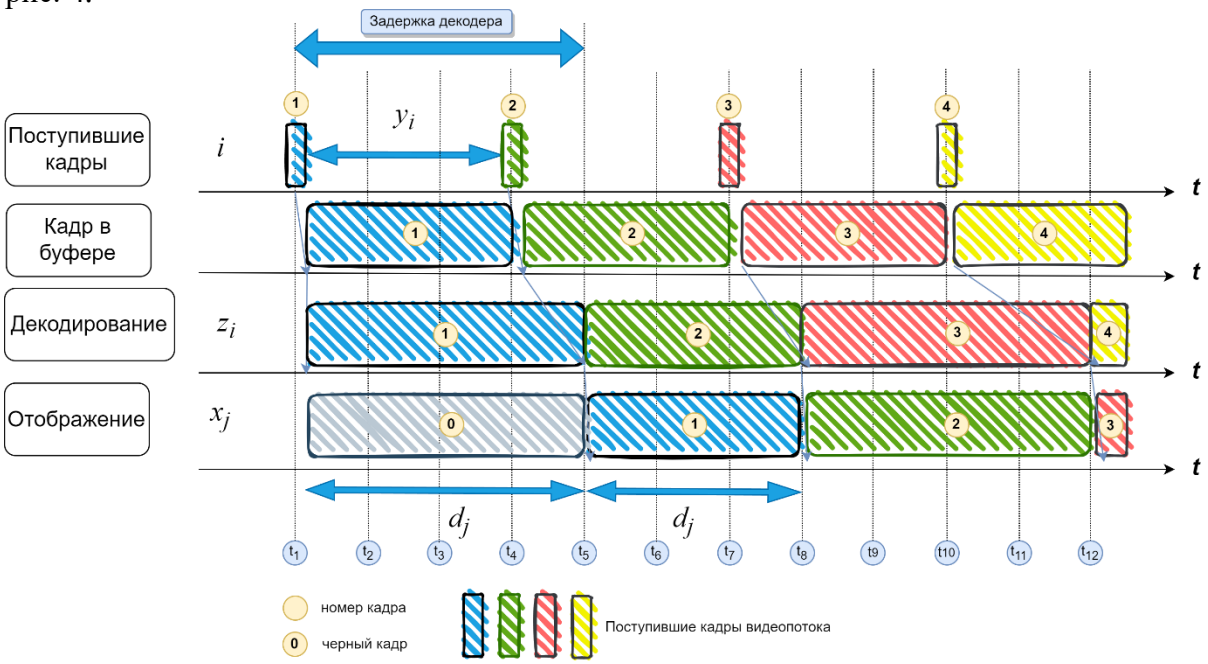


Рис. 4. Процесс последовательного отображения кадров видеопотока

Данная модель обработки кадров на стороне декодера справедлива в случае последовательного выполнения этапов декодирования на стороне СВП и не учитывает наличие предсказательной модели. В случае, когда отбор кадров для отображения на стороне ВП выполняется параллельно с заданным FPS, актуальные кадры в моменты времени j и d_i получаются в соответствии с выражениями:

$$x_j : j = \max(j) : \begin{cases} \left(\sum_{i=0}^{j} y_i = \left(\sum_{k \in J_t} z_k \right) - z_j \right) \\ t' = t - d_i \times \left\lfloor \frac{t}{d_i} \right\rfloor \end{cases},$$

$$x_j : j = \min(j) : \begin{cases} \max(j) : \left(\sum_{i=0}^{j} y_i = t_j \right), t_j \le t' \\ t' = t - d_i \times \left\lfloor \frac{t}{d_i} \right\rfloor \end{cases},$$

$$d_i = \text{const} = \frac{1}{\text{FPS}} (\text{c}) .$$

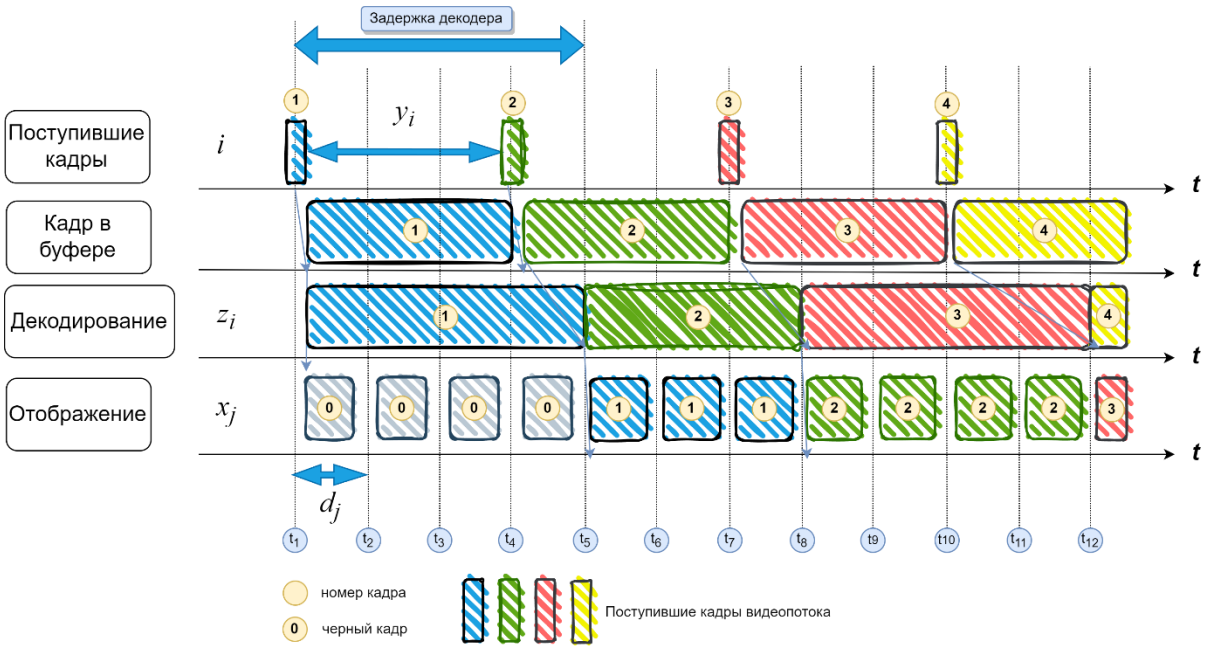


Рис. 5. Процесс параллельного отображения кадров видеопотока

Следует отметить, что свести задачу предсказания кадров FPV-видеопотока к задаче интерполяции промежуточных кадров только из кадров, поступивших из канала связи, невозможно [7]. В момент i на СВП имеется только кадр x_i, а также предыдущие промежуточные кадры x_{i-1}', x_{i-2}', x_{i-3}' и далее. Для интерполяции кадров x_{i+1}', x_{i+2}', x_{i+3}' требуется кадр x_{i+d}, но в момент времени i, когда требуется выполнить интерполяцию, на СВП его нет. В этой связи необходимо использовать нейросетевые алгоритмы, позволяющие предсказывать необходимые кадры на основе предыдущих.

В данной работе рассматривается частный случай предсказания n кадров на основе $n+1$ предыдущих кадров, включающих в том числе и один реальный. Это обусловлено высокой динамикой сцены при FPV-управлении БПЛА.

Однако после поступления в очередь первого кадра x_0 в ней отсутствуют необходимые n предыдущих предсказанных кадров. Соответственно, выполнение предсказания по предложенной схеме в начальный момент времени невозможно. Требуется некоторое время на запуск процесса предсказания, который может быть выполнен несколькими способами:

1. Предсказание с первых кадров. Получение первых n кадров из канала связи и запуск процесса предсказания из кадров x_0, x_1, x_2… x_n. Минусом данного подхода являются различные временные промежутки между этими кадрами и кадрами для предсказания, которые будут сглаживаться по мере продолжения процесса предсказания.

2. Первичная интерполяция. Получение первых двух кадров x_0 и x_1, интерполяция n–1 кадров между ними и запуск процесса предсказания. Минусом данного подхода является пропуск многих деталей при достаточно больших n и y_0.

3. Пошаговое предсказание. Пусть h_0, h_1, h_2, …, h_k – неубывающая последовательность шагов предсказания, такая что $h_k = n$. Сперва выполняется получение h_0+1 первых кадров и предсказывается h_0 промежуточных кадров. Далее после поступления дополнительных кадров из h_1+1 кадров предсказываются (из всего объема кадров) h_1 кадров и так далее до достижения n предсказываемых кадров. Метод является модификацией предсказания с первых кадров с некоторым сглаживанием различных временных промежутков. Для одной из наиболее простых модификаций предложенного алгоритма необходимо n шагов для начала процесса предсказания: $h_0 = 1$, $h_{i+1} = h_i + 1$.

4. Распределённая интерполяция. Пусть для интерполяции используется k кадров. Тогда между x_i и x_{i+1} кадрами интерполируется n_i кадров таким образом, что числа интерполируе-

мых кадров удовлетворяют выражению $\sum_{i=0}^{k}(n_i-1)-1=n$, если первый кадр входит в исходные для следующего предсказания, и выражению $\sum_{i=0}^{k}(n_i-1)=n$, если не входит. Данный метод является обобщённым случаем первичной интерполяции и позволяет уменьшить пропуск многих деталей при достаточно высоких n и y_i. Из него можно вывести ряд методик. Например, первый кадр в предсказании не участвует, а все n_i взяты равными, тогда k определяется по формуле

$$k=\frac{n}{n_i-1}$$

В настоящей работе используется наиболее простой первый способ запуска процесса предсказания. Способ запуска процесса важен только с точки зрения качества восстанавливаемых кадров, а не времени выполнения.

4. Методика проведения эксперимента

В ходе экспериментов определяются параметры нейросетевого декодера при поступлении 1000 кадров из авторского набора данных и измеряются целевые метрики. Поступление следующего кадра осуществляется сразу после завершения работы всех блоков по обработке предыдущего кадра.

Рис. 6. Пример кадров из авторского набора данных

Параметр разрешения исходных кадров $a \times b$ (a – ширина кадра в пикселях, b – высота кадра в пикселях) неизменен и равен 512×512 [11].

Число предсказываемых кадров n варьируется от 0 до 5. Согласно исследованию [6] уже на пятом предсказываемом кадре начинают проявляться видимые дефекты, поэтому предсказывать большее число кадров нецелесообразно из-за значительного ухудшения качества.

Выходное разрешение $w \times h$ принимает следующие значения: 512×512, 1280×720 (HD) и 2560×1440 (QHD) (размерность FullHD 1920×1080 не поддерживается моделью DMVFN). При разрешении HD число пикселей приблизительно в 3.52 раза больше, чем в исходном разрешении, а при разрешении QHD – в 14 раз. Пикселей в QHD больше, чем в HD, в 4 раза. Данных показателей достаточно для проверки линейности времени предсказания кадров моделью. Итого проводится 18 экспериментов.

Целевые метрики отражают время выполнения отдельных блоков нейросетевого декодера: время декомпрессии t_{DC}, время декодирования t_{DCD}, время интерполяции t_I и время предсказания t_P. На их основе можно определить полное время работы декодера t_n:

$$t_n = t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + t_P \left(\text{мс}\right). \tag{2}$$

Время декомпрессии t_{DC} также включает время выполнения операций чтения буфера и перевода данных во внутренний формат системы. Параметр t_I включает в себя в том числе время перевода внутреннего формата данных из torch.tensor в numpy.ndarray. Метрики времени работы измеряются в миллисекундах (мс). Частоту кадров FPS можно определить в соответствии с выражением:

$$\mathrm{FPS} = \frac{\left(n+1\right)\times 1000}{t_n}\left(\text{к/с}\right). \tag{3}$$

Время работы очереди кадров не учитывается из-за его незначительности. Используемая доверительная вероятность равна 95 %.

Следует отметить, что операции на GPU с помощью CUDA выполняются асинхронно, что не позволяет определить время их выполнения, а значит, и связанные метрики точно. Попытка измерения времени выполнения декодера приводит к получению достаточно точного общего времени работы с одним кадром, но некорректного времени выполнения отдельных блоков. Тем не менее способ приближённого вычисления времени выполнения каждого из блоков заключается в использовании синхронизации после выполнения каждого блока (torch.cuda.syncronize). Это приводит к некоторому увеличению общего времени работы, но позволяет корректно измерять целевые метрики.

После получения первичных результатов проводится их анализ. Получаются зависимости времени предсказания t_P от числа предсказываемых кадров n и числа пикселей N_p всех предсказываемых кадров $N_p = \mathrm{width} \times \mathrm{height} \times n\left(\text{пикселей}\right)$, где N_p для исходного разрешения равно 262144×n, для HD – 921600×n, а QHD – 3686400×n.

На основе данных о зависимости числа пикселей N_p и времени предсказания t_P методом наименьших квадратов строится регрессионная модель. В случае линейной зависимости она определяется формулой:

$$\hat{t}_P = b_0 + b_1 \times N_p = b_1 \times N_p \left(\text{мс}\right),$$

где $b_0 = 0$ и b_1 – коэффициенты регрессии.

Вычисляется интервальная оценка [12] в соответствии с выражением:

$$\hat{t}_P \in \left[b_1 \times N_p \pm t_{a,n-2} \times \hat{S} \times \sqrt{\frac{1}{n} \times \left(1 + \frac{\left(\bar{N}_p - N_p\right)^2}{Var\left(N_p\right)} \right)} \right], \tag{4}$$

где $t_{a,n\text{-}2}$ – коэффициент Стьюдента с уровнем значимости a и n–2 степенями свободы, $\hat{S}$ – стандартная ошибка остатков (стандартная ошибка оценки или регрессии), которая вычисляется по формуле:

$$\hat{S} = \sqrt{\frac{\sum\left(t_P - \hat{t}_P\right)^2}{n-2}}$$ [12, с. 53].

Модель обобщается на другое оборудование путём оценки зависимости и ввода дополнительного коэффициента. Оценивается, является ли она гомоскедастичной и какова её область применения.

Полученная модель используется для нахождения такой конфигурации блоков нейросетевого декодера, при которой максимизируется FPS, а также границ данных конфигураций. Во-первых, рассчитывается минимальный период прихода кадров, при котором модель способна обеспечить увеличение FPS ($T_{\min}$). Во-вторых, оценивается максимальное FPS_{max} данного нейросетевого кодека и условия его достижения. В-третьих, проверяется правильность порядка блоков в нейросетевом декодере для максимизации величины FPS.

5. Эксперименты

Результаты экспериментов по заданным исходным данным и средние значения целевых метрик с доверительными интервалами представлены в табл. 1.

Таблица 1. Результаты экспериментов

n, шт.	w x h, пикс.	t_{DC}, мс	t_{DCD}, мс	t_I, мс	t_P, мс	t_n, мс	FPS, к/с
0	512×512	0,001 ±0,002	26.01 ±0.028	1.08 ±0.0169	0	27.1 ±0.036	36.92 ±0.03
0	1280×720	0	26	1.13 ±0.02	0	27.13 ±0.02	36.87 ±0.027
0	2560×1440	0	26	1.12 ±0.021	0	27.12 ±0.021	36.88 ±0.028
1	512×512	0	26.41 ±0.03	1.02 ±0.008	11.76 ±0.109	39.18 ±0.113	51.12 ±0.12
1	1280×720	0	26 ±0.004	1.89 ±0.049	29.92 ±0.16	57.82 ±0.18	34.67 ±0.106
1	2560×1440	0	26.53 ±0.04	2.02 ±0.098	175.88 ±0.45	204.44 ±0.453	9.79 ±0.022
2	512×512	0	26 ±0.003	1.05 ±0.014	22.05 ±0.089	49.11 ±0.091	61.14 ±0.114
2	1280×720	0	26	1.58 ±0.03	54.52 ±0.152	82.1 ±0.155	36.57 ±0.07
2	2560×1440	0	26.57 ±0.037	2.02 ±0.01	325.17 ±0.56	353.76 ±0.563	8.48 ±0.014
3	512×512	0	26.23 ±0.026	1.06 ±0.015	32.95 ±0.195	60.24 ±0.217	66.62 ±0.231
3	1280×720	0	26.36 ±0.03	1.76 ±0.228	76.28 ±0.228	104.39 ±0.237	38.37 ±0.088
3	2560×1440	0	27 ±0.004	2.18 ±0.024	433.36 ±2.738	462.55 ±2.748	8.73 ±0.052
4	512×512	0	26.06 ±0.015	1.11 ±0.02	42.84 ±0.179	70.02 ±0.191	71.54 ±0.186
4	1280×720	0	26.65 ±0.03	2.09 ±0.046	104.05 ±0.264	132.79 ±0.284	37.7 ±0.083
4	2560×1440	0	27 ±0.005	2.59 ±0.031	466.38 ±1.691	495.98 ±1.688	10.11 ±0.029

5	512×512	0	26.03 ±0.01	1.17 ±0.023	52.94 ±0.189	80.14 ±0.197	74.98 ±0.177
5	1280×720	0	26 ±0.004	1.83 ±0.026	126.76 ±0.284	154.59 ±0.285	38.84 ±0.072
5	2560×1440	0	26.01 ±0.006	2.04 ±0.014	767.23 ±1.129	795.279 ±1.129	7.55 ±0.012

Несмотря на наличие некоторых флуктуаций средних значений и половин длин доверительных интервалов во времени выполнения декодирования и интерполяции, которые при соответствующих экспериментах выполняются идентично, они приблизительно не превышают одной миллисекунды и не могут рассматриваться как существенные. Таким образом, для проведения экспериментов по определению времени выполнения на производительном оборудовании желательно использовать б*о*льший набор данных и проводить измерения уже в наносекундном диапазоне. Однако полученные результаты (табл. 1) достаточно точны для выполнения цели и задач настоящей работы.

Кроме этого, получены зависимости среднего времени предсказания t_P от числа предсказываемых кадров n (рис. 7) и корреляционная зависимость времени предсказания t_P числа пикселей N_p (рис. 8). Данные зависимости близки к линейным. В настоящей работе для предсказания n кадров используется n+1 кадров, следовательно, имеется прямая функциональная зависимость между n и N_p. Соответственно, линейность зависимости средней t_p от n при каждом отдельном разрешении обуславливается линейной корреляционной зависимостью t_p от N_p. На основе всех 18000 проведённых испытаний для этих двух величин был рассчитан коэффициент корреляции Пирсона: 0.9838, что близко к функциональной зависимости.

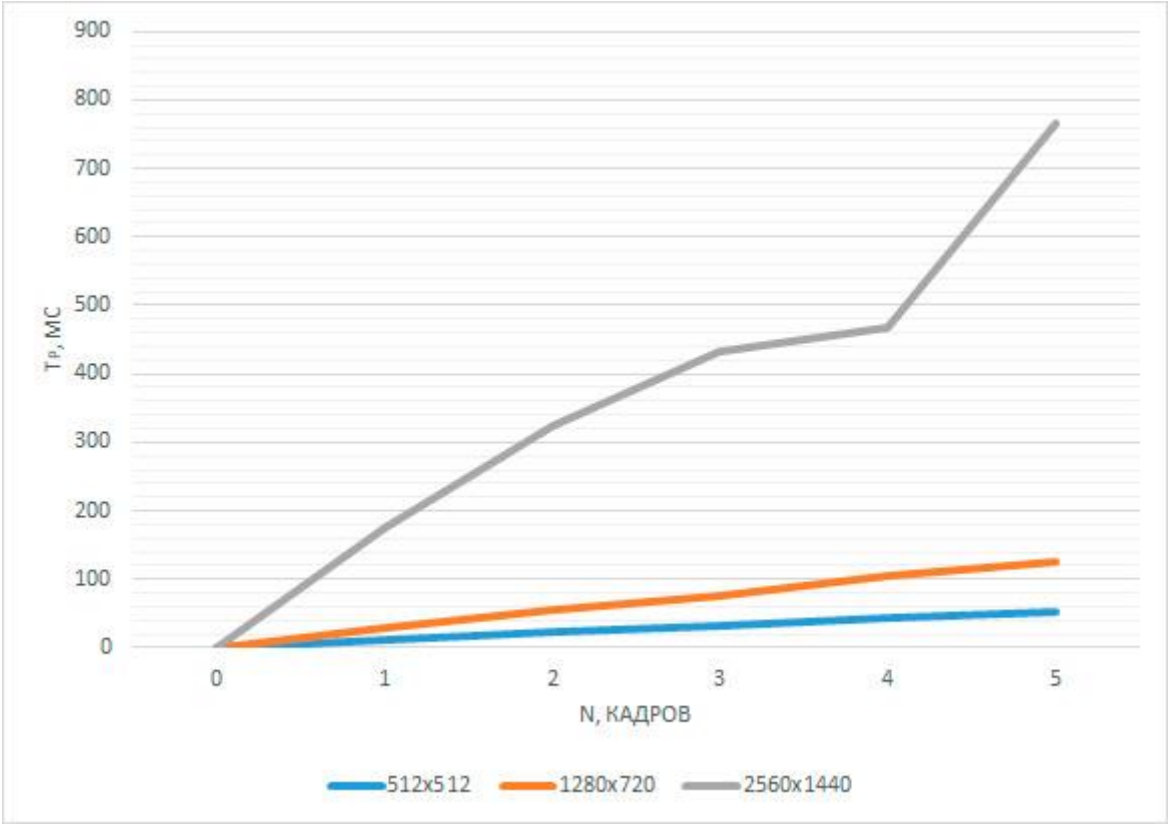


Рис. 7. Зависимость среднего времени предсказания от числа предсказанных кадров

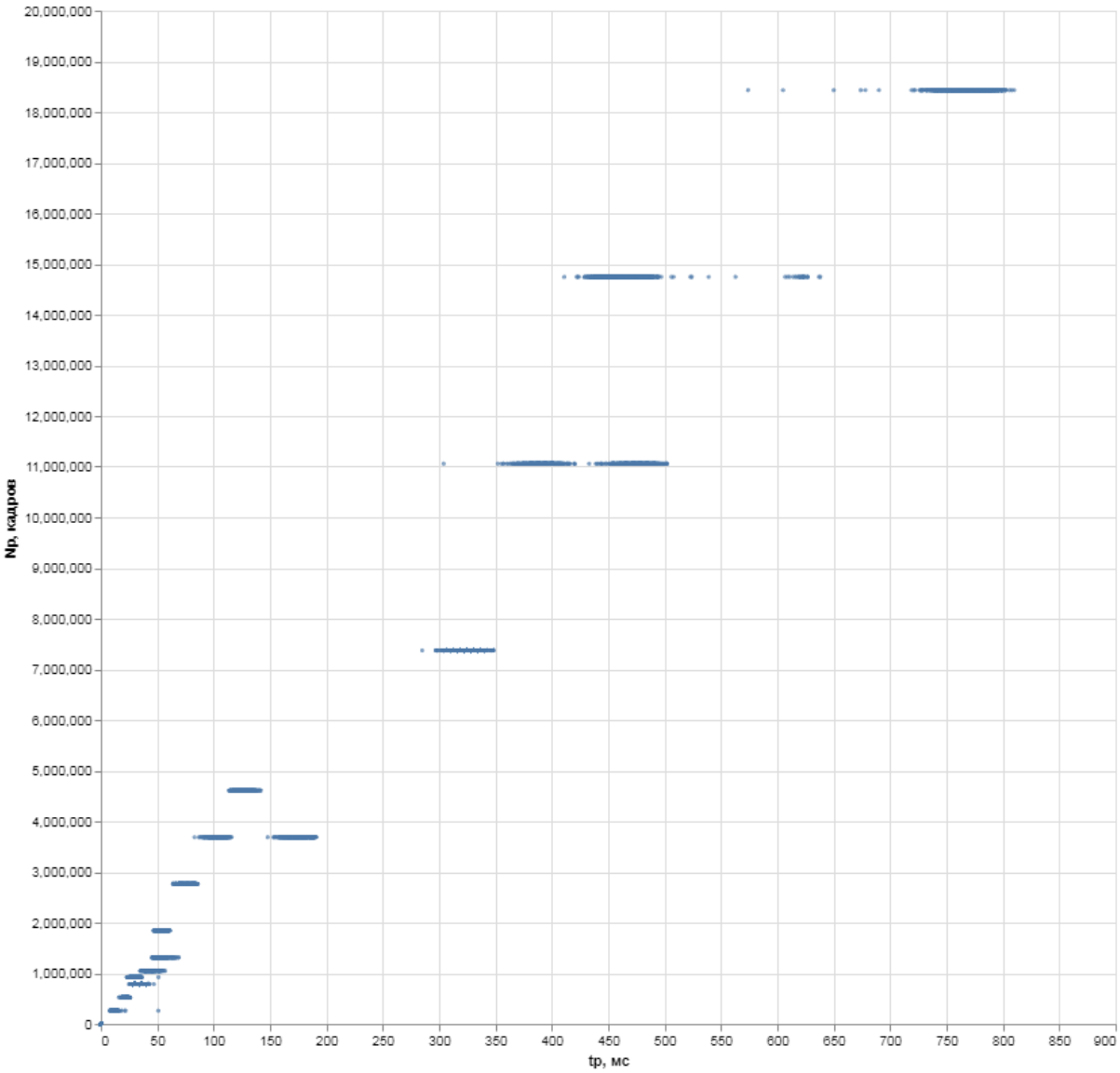


Рис. 8. Корреляционная зависимость времени предсказания от числа пикселей кадров

Отметим, что влияние отдельных компонент $N_{\text{полн}}$:

$$N_{\text{полн}} = N_p + N_{sp} = \text{height} \times \text{width} \times \left(n_s + n_d\right)\left(\text{пикселей}\right),$$

где n_s – число исходных кадров, n_d – число предсказываемых кадров, N_{sp} – число пикселей исходных кадров, а $N_{\text{полн}}$ – число пикселей всех кадров, требует отдельного исследования. В данной работе N_{sp}, а значит, и $N_{\text{полн}}$ функционально зависят от N_p, что не позволяет выполнить полноценное исследование влияния отдельных компонент на t_p. Отдельного рассмотрения также требует большое число предсказываемых кадров n. Предполагается, что по мере исчерпания памяти видеокарты зависимость между t_p и N_p перестанет быть линейной и будет ближе к степенной при степени, лежащей в интервале (0; 1).

6. Прогнозирование времени предсказания кадров

Построена линейная регрессионная модель для предсказания времени предсказания кадров от числа пикселей исходных кадров при выполнении неоптимизированной нейронной сети DMVFN на описанном ранее оборудовании $\hat{t}_P = 0.0000378 \times N_p\left(\text{мс}\right)$.

При доверительной вероятности, равной 95 %, числе испытаний 18000 (по 1000 в каждом эксперименте) на основе (4) вычислена интервальная оценка прогнозируемого времени предсказания t_p (рис. 9):

$$\hat{t}_P \in \left[0.0000378 \times N_p \pm 1767.15255 \times \sqrt{\frac{1}{17998} \times \left(1 + \frac{\left(4058453.33333 - N_p\right)^2}{27857767920071.113} \right)} \right]. \quad (5)$$

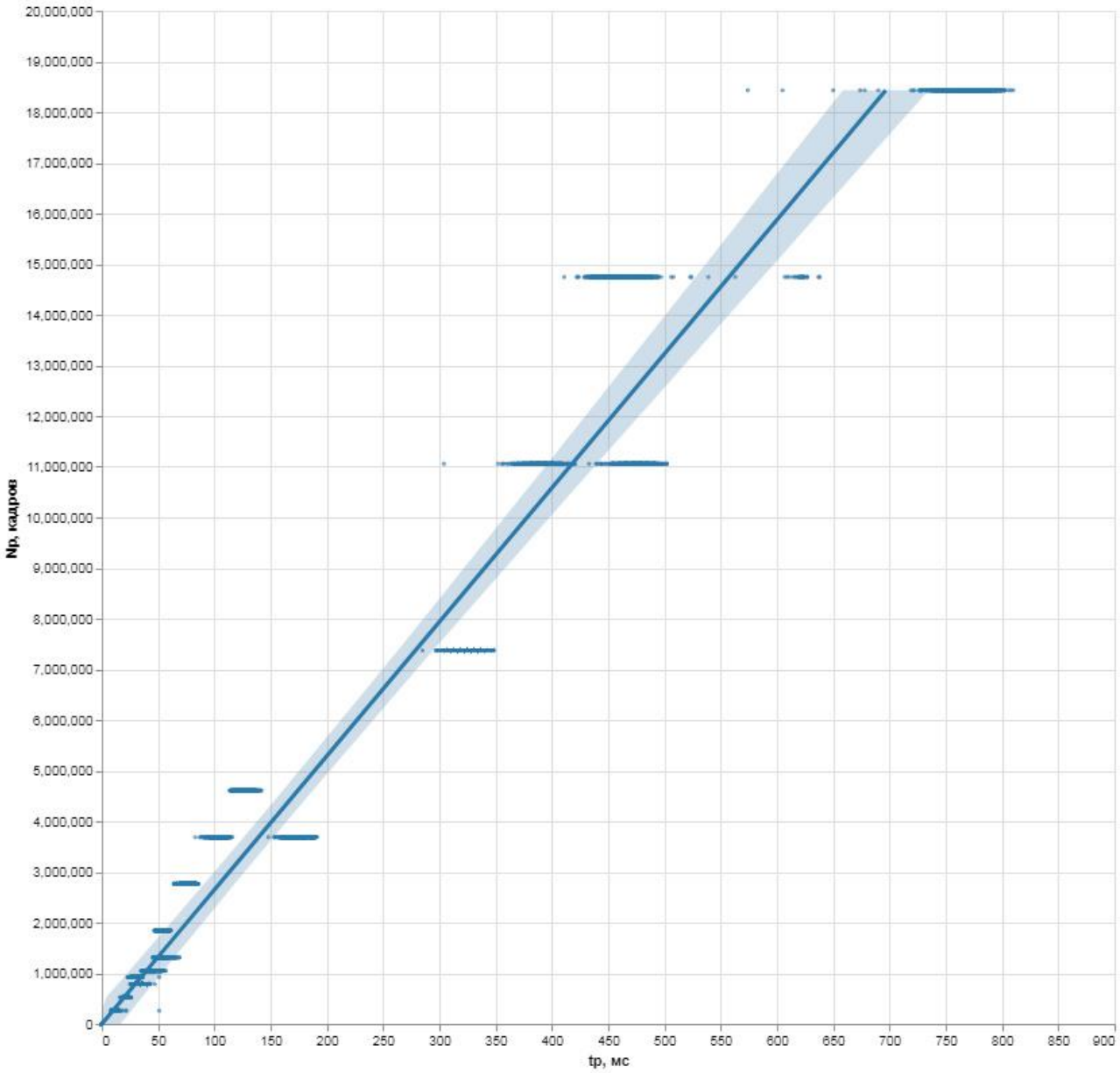


Рис. 9. Интервальная оценка предложенной регрессионной модели

Как можно видеть из рис. 9, модель гетероскедастична: в то время как математическое ожидание ошибки близко к нулю, дисперсия ошибки модели возрастает при возрастании N_p. Соответственно, оценки, получаемые с помощью данной модели, при достаточно высоких N_p неэффективны. Тем не менее в рамках текущей работы данная модель используется в зоне достаточно малой дисперсии ошибки (её области компетенции). В качестве альтернативы возможно увеличение интервала оценок при достаточно больших N_p.

Выдвинем гипотезу, требующую для подтверждения отдельного исследования. Пусть n насколько мало, что при выполнении нейронной сети DMVFN на CUDA видеопамять видеокарты далека от исчерпания. Тогда линейный характер корреляционной зависимости ве-

личин t_p и N_p сохраняется. Соответственно, приведённая выше модель может быть актуализирована и для другого оборудования.

Для актуализации регрессионной модели не обязательно повторять все проведённые эксперименты с нуля. Достаточно получить среднее время выполнения t'_{pcp} при некотором множестве N предсказываемых кадров с суммарным числом пикселей N_p. Тогда регрессионная модель при достаточно малых n может в общем случае быть представлена как:

$$\hat{t}_P = \frac{0.0000378 \times N_p}{k_p} (\text{мс}), \tag{6}$$

где k_p показывает, во сколько раз быстрее нейронная сеть выполняется на видеокарте A100, чем на тестируемой видеокарте, и вычисляется по формуле

$$k_p = \frac{t_{\text{pcp}}}{t'_{\text{pcp}}}$$

при

$$t'_{\text{pcp}} = \frac{\sum_{t'_p : N_p \in N} (t_p)}{x}.$$

где x – число испытаний на тестируемой видеокарте.

В свою очередь, t_{pcp} для нескольких N_p можно получить как среднее арифметическое значение при каждом конкретном значении N_p, т.к. в данной работе в каждом эксперименте – равное число испытаний (1000 шт.), а каждому N_p соответствует свой эксперимент. Обозначим данную регрессионную модель как *модель DMVFN малых n*. Соответственно, разновидности этой модели для различных тестируемых видеокарт – *частными случаями модели DMVFN малых n*.

Пусть k_p достаточно точно отражает отношение t_p к t'_p для всех N_p, на которых построена исходная регрессионная модель. Тогда интервал значений для модели DMVFN малых n элементарно выводится из формул (4), (5), (6) и может быть приблизительно рассчитан как

$$\hat{t}_P \in \left[\frac{0.0000378 \times N_p}{k_p} \pm \sqrt{\frac{1378.8659^2 \times 17998 + \sum \left(t'_p - \hat{t}_p \right)^2}{17998 + x}} \times t_{0,05,17998+x} \times \right.$$

$$\left. \times \sqrt{\frac{1}{17998 + x} \cdot \left(1 + \frac{(17999 + x) \times \left(\frac{4058453.33333 \times 18000 + \sum N'_p}{18000 + x} - N_p \right)^2}{27857767920071.113 \times 17999 + \sum \left(N'_p - \frac{4058453.33333 \times 18000 + \sum N'_p}{18000 + x} \right)} \right)} \right] (\text{мс}),$$

где N'_p – это N_p из испытаний на тестовой видеокарте, для которой находится частный случай модели DMVFN малых n. Данная формула основана на пересчёте числа степеней свободы, стандартной ошибки оценки, среднего значения и дисперсии.

7. Оценка условий увеличения FPS блоком предсказания

На основании (2) и (3) задачу максимизации FPS для текущей (рис. 2) конфигурации нейросетевого декодера можно свести к минимизации:

$$\text{FPS} = \frac{(n+1) \times 1000}{t_{\text{DC}} + t_{\text{DCD}} + t_I + t_P + T_з} \left(\frac{\text{кадров}}{\text{с}} \right),$$

где $T_з$ – некоторая дополнительная задержка до поступления кадра из канала связи.

Время выполнения декомпрессии t_{DC} алгоритмом DEFLATE в рамках блока декомпрессии измеряется в наносекундах и несущественно относительно времени выполнения других блоков.

Время декодирования кадра VQ-VAE-f16 в рамках блока декодирования приблизительно одинаково и на видеокарте A100 в среднем составляет 26.3 мс.

Значительные флуктуации средних оценок t_I при одних и тех же операциях блока интерполяции $I(X)$ не позволяют дать корректную оценку с помощью регрессионной модели, как в случае блока предсказания $P(x)$. Для каждого разрешения использованы средние t_I за 6000 испытаний (табл. 1). Таким образом, для разрешения 512×512 t_I = 1.08 мс, для разрешения 1280×720 – 1.71 мс, а для 2560×1440 – 1.99 мс. При этом существенная часть времени выполнения блока интерполяции $I(X)$ – это перевод формата из torch.tensor и numpy.ndarray, а время работы алгоритма бикубической интерполяции ввиду сопоставимости значений обоих показателей и их зависимости от размерности кадра выделить без проведения отдельных измерений не представляется возможным.

Оценим минимальную дополнительную задержку $T_{\min} \geq 0$, при которой возможно увеличение FPS с помощью предиктора. Первым шагом к решению данной задачи является нахождение минимальной дополнительной задержки поступления кадра из канала связи $T_з \geq 0$, такой что неравенство

$$\frac{1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + T_з} < \frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + t_P + T_з},$$

$$\frac{1000}{26.2718 + t_I + T_з} < \frac{(n+1)\times 1000}{26.2718 + t_I + 0.0000378\times N_p + T_з},$$

$$\frac{1000}{26.2718 + t_I + T_з} < \frac{(n+1)\times 1000}{26.2718 + t_I + 0.0000378\times n\times \mathrm{width}\times \mathrm{height} + T_з}$$

имеет решение для некоторых n>0. Для разрешения 512×512:

$$\frac{1000}{26.2718 + 1.0833 + T_з} < \frac{(n+1)\times 1000}{26.2718 + 1.0833 + 0.0000378\times n\times 512\times 512 + T_з},$$

$$\frac{1000}{27.3551 + T_з} < \frac{(n+1)\times 1000}{27.3551 + 9.9090432\times n + T_з}.$$

С учетом

$$\forall T_з : T_з \geq 0 \Rightarrow \forall T_з : (27.3551 + T_з) > 0,$$

$$\forall n, \forall T_з : n > 0, T_з \geq 0 \Rightarrow \forall n, \forall T_з : (27.3551 + 9.9090432\times n + T_з) > 0$$

справедливы следующие неравенства

$$1000\times(27.3551 + 9.9090432\times n + T_з) < (n+1)\times 1000\times(27.3551 + T_з),$$

$$27355.1 + 9909.0432\times n + 1000\times T_з < 27355.1\times n + 1000\times Tз\times n + 27355.1 + 1000\times T_з,$$

$$9909.0432\times n < 27355.1\times n + 1000\times Tз\times n.$$

Следует отметить, что при n = 0 обе стороны неравенства равны, что показывает правильность выполняемых действий. В данном случае предсказание вырождается: предсказание 0 кадров выполняется за 0 секунд, при этом не добавляя кадров, а FPS становится таким же, как и в случае без блока предсказания. Соответственно, $T_з \geq 0$.

Таким образом, при заданных условиях и ограничениях использование модели DMVFN для предсказания изображений разрешения 512×512 не требует дополнительной задержки $T_з$. Это позволяет оценить $T_{\min}$. Если дополнительная задержка $T_з$ > 0, то $T_{\min}$ равен сумме времени обработки нейросетевым декодером предыдущего кадра и дополнительной задержки. Если же $T_{з\,\min}$ = 0, то кадр из канала связи может поступать как в момент завершения обработки декодером предыдущего кадра, так и во время выполнения блоков нейросетевого декодера. Это не оказывает влияния на FPS, т.к. последний кадр буферизируется и поступает

на блок декомпрессии $DC(T^*)$ сразу по завершении обработки нейросетевым декодером предыдущего кадра. Следовательно, в этом случае $T_{\min} = 0$.

Перед анализом остальных разрешений выведем более общую формулу, которая может быть использована в том числе для других видеокарт при достаточно малых n. В случае другого оборудования время выполнения декомпрессии t_{DC} может не быть пренебрежительно мало. По условию

$$\forall T_з, \forall t_I, \forall t_{\text{DC}} : T_з, t_I, t_{\text{DC}} \geq 0, \forall \text{width}, \forall \text{height}, \forall n, \forall t_{\text{DCD}} : \text{width}, \text{height}, n, t_{\text{DCD}} > 0$$

для всех переменных выполняются неравенства

$$\forall T_з, \forall t_I, \forall t_{\text{DC}}, \forall t_{\text{DCD}} : t_{\text{DC}} + t_{\text{DCD}} + t_I + T_з > 0,$$

$$\forall T_з, \forall t_I, \forall t_{\text{DC}}, \forall t_{\text{DCD}}, \forall n, \forall \text{width}, \forall \text{height} :$$

$$t_{\text{DC}} + t_{\text{DCD}} + t_I + \frac{0.0000378 \times n \times \text{width} \times \text{height}}{k_p} + T_з > 0.$$

Следовательно, возможно свести решение

$$\frac{1000}{t_{\text{DC}} + t_{\text{DCD}} + t_I + T_з} < \frac{(n+1) \times 1000}{t_{\text{DC}} + t_{\text{DCD}} + t_I + \frac{0,0000378 \times n \times \text{width} \times \text{height}}{k_p} + T_з}$$

для нахождения допустимых $T_з$ к:

$$1000 \times \left(t_{\text{DC}} + t_{\text{DCD}} + t_I + \frac{0.0000378 \times n \times \text{width} \times \text{height}}{k_p} + T_з \right) <,$$

$$< (n+1) \times 1000 \times (t_{\text{DC}} + t_{\text{DCD}} + t_I + T_з)$$

$$1000 \times t_{\text{DC}} + 1000 \times t_{\text{DCD}} + 1000 \times t_I + 1000 \times T_з +$$

$$+ \frac{0.0378 \times n \times \text{width} \times \text{height}}{k_p} < 1000 \times n \times t_{\text{DC}} + 1000 \times n \times t_{\text{DCD}} + \quad ,$$

$$+ 1000 \times n \times t_I + 1000 \times n \times T_з + 1000 \times t_{\text{DC}} + 1000 \times t_{\text{DCD}} + 1000 \times t_I + 1$$

$$\frac{0.0378 \times n \times \text{width} \times \text{height}}{k_p} < 1000 \times n \times (t_{\text{DC}} + t_{\text{DCD}} + t_I + T_з),$$

$$\frac{0.0000378 \times \text{width} \times \text{height}}{k_p} < t_{\text{DC}} + t_{\text{DCD}} + t_I + T_з.$$

Соответственно, $T_з$ и $T_{\min}$, при которых предикторная модель DMVFN позволяет увеличить FPS нейросетевого декодера, можно оценить как

$$\begin{cases} T_з > \dfrac{0.0000378 \times \text{width} \times \text{height}}{k_p} - t_{\text{DC}} - t_{\text{DCD}} - t_i \\ T_з \geq 0 \end{cases}, \tag{7}$$

$$T_{\min} = \begin{cases} 0, & \dfrac{0.0000378 \times \text{width} \times \text{height}}{k_p} - t_{\text{DC}} - t_{\text{DCD}} - t_i \leq 0 \\ t_{DC} + t_{DCD} + t_I + t_P + T_{з\min}, [...] > 0 \end{cases}. \tag{8}$$

Вторую часть выражения (8) можно значительно упростить:

$$T_{\min} = t_{\text{DC}} + t_{\text{DCD}} + t_I + t_P + \frac{0.0000378 \times \text{width} \times \text{height}}{k_p} - t_{\text{DC}} - t_{\text{DCD}} - t_I,$$

$$T_{\min} = t_P + \frac{0.0000378 \times \text{width} \times \text{height}}{k_p},$$

$$T_{\min} = \frac{(n+1)\times 0.0000378\times \text{width}\times \text{height}}{k_p}.$$

Таким образом, (8) преобразуется в

$$T_{\min} = \begin{cases} 0, & \dfrac{0.0000378\times \text{width}\times \text{height}}{k_p} - t_{\text{DC}} - t_{\text{DCD}} - t_I \le 0 \\ \dfrac{(n+1)\times 0.0000378\times \text{width}\times \text{height}}{k_p}, & \dfrac{0.0000378\times \text{width}\times \text{height}}{k_p} - t_{\text{DC}} - t_{\text{DCD}} - t_I > 0 \end{cases} \quad (9)$$

При определённом условии (9) $T_{\min}$ может быть оценен как время предсказания t_P (n+1) кадра.

Проверка полученных оценок для ранее рассчитанного частного случая видеокарты A100 и разрешения 512×512 показала соответствие ранее полученному результату $T_з \ge 0$. Соответственно, (7) и (9) можно применять для оценки целесообразности использования блока предиктора модели DMVFN при использовании для оценки времени выполнения предиктора t_P модели DMVFN малых n. Так как решение (7) существует и единственно, подтверждается ещё одно утверждение: при выполнении модели DMVFN на любом оборудовании при любой конфигурации нейросетевого декодера обязательно найдётся такое $T_{\min}$, при котором использование блока предсказания $P(x)$ даёт увеличение FPS.

Оценки для остальных разрешений: (1280×720) $T_з > 6{,}8376\times(n+1)$; (2560×1440) $T_з > 139{,}3459\times(n+1)$. Безусловного увеличения FPS не наблюдается.

Полученные результаты хорошо согласуются с экспериментальными данными. Если в случае разрешения 512×512 при использовании DMVFN FPS однозначно повышается, то в случае 1280×720 FPS колеблется, а в случае 2560×1440 – уменьшается и колеблется.

Гипотеза, доказательство которой необходимо для реализации анализатора КИО и его связки с КИВ. Пусть исходный период прихода кадров с кодера $T_к \ge T_{\min}$. Тогда возможно нахождение такого n, что добавление блока предсказания DMVFN в декодер повышает FPS.

Доказательство. *Пусть $T_{min} = 0$, тогда добавление блока предсказания DMVFN безусловно увеличивает FPS. Пусть $T_k > T_{min}$. Тогда, исходя из (9), такие параметры n могут быть найдены как*

$$T_k > \frac{(n+1)\times 0.0000378\times \text{width}\times \text{height}}{k_p},$$

$$n < \frac{T_k \times k_p}{0.0000378\times \text{width}\times \text{height}} - 1.$$

Тем не менее сам факт использования блока предсказания означает, что $n \ge 1$. Поставим условие существования решения

$$\frac{T_k \times k_p}{0.0000378\times \text{width}\times \text{height}} - 1 \ge 1,$$

$$T_k > \frac{2\times 0.0000378\times \text{width}\times \text{height}}{k_p}.$$

Иными словами, задержка должна быть выше, чем время предсказания двух кадров с помощью DMVFN и общее время работы декодера. Исходная гипотеза подтверждается, таким образом, лишь в этом частном случае.

Теорема о целесообразности блока предсказания КИВ. *Если период y_i поступления кадров из канала связи на декодер больше общего времени работы декодера и больше или равен времени предсказания двух кадров нейронной сетью DMVFN, существует такое число предсказываемых кадров n, что для любых разрешений кадров добавление блока предсказания DMVFN в нейросетевой декодер обеспечивает увеличение FPS.*

Первая задача работы, которая заключается в оценке целесообразности использования блока DMVFN, выполнена.

8. Анализ максимального FPS

Оценим максимальное FPS (FPS_{max}) для каждого из разрешений: 512×512, 1280×720 и 2560×1440. Справедливо выражение

$$\mathrm{FPS} = \frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + t_P + T_з}\left(\frac{\text{кадров}}{\text{с}}\right). \tag{10}$$

Преобразуем (10) в

$$\mathrm{FPS} = \frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + \dfrac{0.0000378\times n\times \mathrm{width}\times \mathrm{height}}{k_p} + T_з}\left(\frac{\text{кадров}}{\text{с}}\right).$$

Если t_{DC}, t_{DCD}, t_I, $T_з$, width и height обязательны при увеличении уменьшают FPS, то интерес представляет последний параметр – число предсказываемых кадров n. Возьмём производную FPS по n:

$$\frac{\delta \mathrm{FPS}}{\delta n} = \delta\left(\frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + \dfrac{0.0000378\times n\times \mathrm{width}\times \mathrm{height}}{k_p} + T_з}\right) \Big/ \delta n,$$

$$\frac{\delta \mathrm{FPS}}{\delta n} = \delta\left(\frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + \dfrac{0.0000378\times n\times \mathrm{width}\times \mathrm{height}}{k_p} + T_з}\right) \Big/ \delta n,$$

$$\frac{\delta \mathrm{FPS}}{\delta n} = \frac{1000\times\left(t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + \dfrac{0.0000378\times n\times \mathrm{width}\times \mathrm{height}}{k_p} + T_з\right)}{\left(t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + \dfrac{0.0000378\times n\times \mathrm{width}\times \mathrm{height}}{k_p} + T_з\right)^2} -$$

$$- \frac{\dfrac{0.0378\times n\times \mathrm{width}\times \mathrm{height}}{k_p} + \dfrac{0,0378\times \mathrm{width}\times \mathrm{height}}{k_p}}{\left(t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + \dfrac{0.0000378\times n\times \mathrm{width}\times \mathrm{height}}{k_p} + T_з\right)^2},$$

$$\frac{\delta \mathrm{FPS}}{\delta n} = \frac{1000\times\left(t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + T_з\right) - \dfrac{0.0378\times \mathrm{width}\times \mathrm{height}}{k_p}}{\left(t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + \dfrac{0.0000378\times n\times \mathrm{width}\times \mathrm{height}}{k_p} + T_з\right)^2}.$$

Производная функции FPS от n имеет кратный разрыв 2-го рода в точке

$$t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + \frac{0.0000378\times n\times \mathrm{width}\times \mathrm{height}}{k_p} + T_з = 0,$$

$$n = -\frac{k_p \times \left(t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + T_з\right)}{0.0000378 \times n \times \mathrm{width} \times \mathrm{height}}. \quad (11)$$

Тем не менее, так как $n > 0$, что вследствие дискретности n тождественно $n \geq 1$, а кратная точка разрыва (11) имеет значение меньше нуля, она выпадает из рассмотрения. Соответственно, при $n \geq 1$ FPS является гладкой функций и знак производной FPS от n одинаков. Существует три случая задачи максимизации FPS: 1) когда по мере увеличения n он убывает, 2) когда по мере увеличения n он возрастает и 3) когда он статичен. Рассмотрим условие первого случая:

$$1000 \times \left(t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + T_з\right) - \frac{0.0378 \times \mathrm{width} \times \mathrm{height}}{k_p} < 0,$$

$$t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + T_з < \frac{0.0000378 \times \mathrm{width} \times \mathrm{height}}{k_p}.$$

Его можно сформулировать следующим образом: время предсказания одного кадра превышает время выполнения всех остальных операций декодера. В этом случае $\mathrm{FPS}_{\max}$ достигается минимизацией n, то есть предсказанием исключительно одного кадра.

Условие второго случая обратно первому

$$t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + T_з > \frac{0.0000378 \times \mathrm{width} \times \mathrm{height}}{k_p}. \quad (12)$$

Время предсказания одного кадра меньше времени выполнения всех остальных операций декодера. В этом случае определим *теоретический предельный FPS*. Он достигается при стремлении n к бесконечности, и его фактическое достижение требует бесконечной видеопамяти используемой видеокарты, что практически невозможно.

Условие третьего случая представляется выражением

$$t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_I + T_з = \frac{0.0000378 \times \mathrm{width} \times \mathrm{height}}{k_p}.$$

Если число предсказываемых кадров n достаточно мало, так что при выполнении DMVFN на CUDA видеопамять используемой видеокарты далека от исчерпания, то для всех таких n FPS неизменен. Таким образом, для экономии видеопамяти следует предсказывать в данном случае только 1 кадр.

$\mathrm{FPS}_{\max}$ для разрешений 512×512, 1280×720 и 2560×1440 при использовании видеокарты A100 соответственно равно: 100.9179, 31.8365 и 11.9323 к/с.

Полученные результаты в целом соотносятся с экспериментальными данными. В то время как при увеличении n при 512×512 наблюдалось стойкое увеличение FPS, при 1280×720 были некоторые флуктуации как с увеличением, так и с уменьшением, а при 2560×1440 – явное понижение FPS.

В то же время результаты показывают ограниченность модели. Согласно экспериментам, $\mathrm{FPS}_{\max}$ (при $n = 1$) для 1280×720 составил 34.6748 к/с, а для 2560×1440 – 9.7956 к/с, при этом при больших n для данных разрешений удалось достичь незначительно больших FPS. Данные особенности объясняются ограниченностью модели и сильными флуктуациями FPS в процессе измерений. Тем не менее можно утверждать, что предложенный метод применим при необходимости достаточно грубой оценки и проектировании систем анализатора КИО и модулей КИВ.

9. Анализ конфигурации блоков нейросетевого декодера

Процесс интерполяции и прогнозирования рассматриваемыми в настоящей работе моделями невозможно выполнить в латентном пространстве признаков, соответственно, данные блоки должны располагаться после блока декодирования *DCD*(*T*). Декомпрессию нельзя

проводить после декодирования, так как для декодирования необходимо наличие уже декомпрессированных данных. Следовательно, блок декомпрессии $DC(T^*)$ должен быть расположен до блока декодирования. С учетом этого возможна только одна альтернативная конфигурация нейросетевого декодера: прогнозирование кадров с последующей их интерполяцией. FPS для такой конфигурации оценивается как (рис. 10):

$$\mathrm{FPS} = \frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_{P(512\times512)} + (n+1)\times t_I + T_з}\left(\frac{\text{кадров}}{\text{с}}\right).$$

Данный вариант обладает следующими преимуществами: значительное уменьшение времени прогнозирования, возможность полноценного увеличения FPS нейросетевого декодера (см. предыдущий раздел) и вместе с тем предоставление кадров видеопотока оператору БВС в нужном разрешении. При этом есть и недостаток: при использовании для увеличения разрешения кадров моделей нейронных сетей сверхразрешения [13–14] возможно значительное уменьшение FPS.

Рис. 10. Альтернативный нейросетевой декодер

Данная конфигурация нейросетевого декодера (рис. 10) предпочтительней исходной (рис. 2) при определённых условиях:

$$\frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_{P(512\times512)} + (n+1)\times t_I + T_з} > \frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + t_P + t_I + T_з},$$

$$\frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + n\times\dfrac{0.0000378\times512\times512}{k_p} + (n+1)\times t_I + T_з} >$$

$$> \frac{(n+1)\times 1000}{t_{\mathrm{DC}} + t_{\mathrm{DCD}} + n\times\dfrac{0.0000378\times \mathrm{width}\times \mathrm{height}}{k_p} + t_I + T_з},$$

$$t_{\mathrm{DC}} + t_{\mathrm{DCD}} + n\times\frac{0.0000378\times \mathrm{width}\times \mathrm{height}}{k_p} + t_I + T_з >$$

$$> t_{\mathrm{DC}} + t_{\mathrm{DCD}} + n\times\frac{0.0000378\times512\times512}{k_p} + (n+1)\times t_I + T_з,$$

$$n\times\frac{0.0000378\times \mathrm{width}\times \mathrm{height}}{k_p} + t_I > n\times\frac{0.0000378\times512\times512}{k_p} + n\times t_I + t_I,$$

$$n \times \frac{0.0000378 \times \text{width} \times \text{height}}{k_p} > n \times \frac{0.0000378 \times 512 \times 512}{k_p} + n \times t_I$$

$$t_I < \frac{1}{k_p} \times \left(0.0000378 \times \text{width} \times \text{height} - 9.9090432\right).$$

Под предпочтительностью понимается более высокое FPS.

В результате получено ограничение на время интерполяции одного кадра, при котором альтернативная модель нейросетевого декодера является предпочтительной. При использовании бикубической интерполяции данное условие выполняется при разрешении, которое несколько больше чем 512×512, в том числе для разрешений 1280×720 и 2560×1440.

Тем не менее переход к альтернативной архитектуре нейросетевого декодера не позволяет увеличить FPS при разрешении кадров видеопотока 512×512. Так как время бикубической интерполяции t_I очень мало по сравнению со временем прогнозирования t_P, случай нахождения FPS_{max} тот же, что и у прогнозирования кадров разрешения 512x512 – второй (при увеличении n FPS увеличивается (12)). Через предел n, стремящегося к бесконечности, применительно к функции FPS для альтернативной архитектуры нейросетевого декодера, рассчитывается FPS_{max} для разрешения 1280×720 – 86.04 к/с и для 2560×1440 – 84 к/с.

В случае, если время интерполяции сопоставимо со временем предсказания, альтернативная архитектура дает результаты, сравнимые с исходной. Соответственно, достижение большего FPS в этом случае связано с оптимизацией и выбором альтернативных моделей интерполяции и прогнозирования, что, вероятнее всего, приведёт к ухудшению качества изображения, подаваемого оператору БВС. В этой связи перед использованием тех или иных моделей в рамках нейросетевого декодера должен быть проведен анализ FPS_{max} и условий достижения данного значения FPS.

10. Заключение

Полученные результаты показывают возможность применения блока прогнозирования для увеличения FPS на стороне оператора БВС. Экспериментально наибольший FPS в 74.98 кадров в секунду был достигнут при выходном разрешении 512×512, модели предсказания DMVFN и предсказании 5 кадров. При выходном разрешении теоретический предел увеличения FPS составляет 100.92 кадров в секунду, к которому можно приблизиться при увеличении числа предсказываемых кадров. Кроме того, при использовании предложенной альтернативной архитектуры теоретический предел при выходном разрешении 1280×720 составляет 86.04 кадров в секунду, а при выходном разрешении 2560×1440 – 84 кадра в секунду.

На основе предложенной регрессионной модели времени предсказания, модели DMVFN малых n, в настоящей работе был получен ряд результатов.

1. Время декомпрессии алгоритмом LZMA в рамках нейросетевого декодера является крайне малым и не оказывает заметного влияния на время обработки поступающего с кодера кадра нейросетевым декодером.

2. При достаточно малом числе предсказываемых кадров зависимость между временем предсказания и числом предсказываемых кадров близка к линейной, что обуславливается линейной зависимостью между числом пикселей предсказываемых изображений.

3. При разрешении кадров видеопотока 512x512 использование модели DMVFN безусловно приводит к увеличению FPS на нейросетевом декодере. В то же время при разрешениях кадров 1280x720 и 2560x1440 увеличение FPS возможно при задержке прихода кадров с кодера, превышающей время работы декодера. При отсутствии таковой задействование модели приводит к колебанию либо уменьшению FPS.

4. Всегда найдётся такая достаточно высокая задержка прихода кадров с кодера, что использование модели DMVFN позволит увеличить FPS.

5. Если исходный период прихода кадров с кодера больше общего времени работы декодера и больше или равен времени предсказания двух кадров нейронной сетью DMVFN, тогда для любых разрешений кадров возможно добавление в декодер блока предсказания DMVFN с таким числом предсказываемых кадров, что FPS на стороне оператора БВС повысится. Для некоторых отдельных разрешений кадров, таких как 512x512, добавление блока предсказания позволяет безусловно увеличивать FPS.

6. Если время предсказания одного кадра нейронной сетью DMVFN больше или равно сумме времён выполнения остальных операций нейросетевого декодера, стоит предсказывать только один кадр. В противном случае – наибольшее количество, которое позволяет видеопамять используемой для работы нейронной сети видеокарты.

7. Если используется интерполяция кадров видеопотока, то при достаточно малом времени её выполнения можно увеличить FPS нейросетевого декодера за счёт перестановки местами блоков интерполяции и прогнозирования.

Предложенные в настоящей работе модели и методы формализуют проблему использования моделей предсказания в нейросетевых декодировщиках и могут служить в качестве методической основы для проектирования соответствующих систем.

Возможны несколько направлений развития исследования.

1. Исследование влияния числа пикселей исходных и предсказываемых кадров на время предсказания.

2. Проведение испытаний на ином оборудовании.

3. Проведение испытаний с б*о*льшим числом предсказываемых кадров.

4. Рассмотрение возможности интерполяции и предсказания кадров видеопотока в латентном пространстве.

5. Проверка и уточнение модели прогнозирования времени предсказания кадров с помощью DMVFN для других видеокарт.

6. Построение модели нейросетевого декодера с учётом блока квантования.

7. Проведение испытаний с набором данных достаточно большого размера, чтобы существенные флуктуации в измеряемых временных показателях и FPS в существенной степени сглаживались.

8. Построение гомоскедастичной регрессионной модели прогнозирования времени предсказания.

Литература

1. *Чмелев В. С., Калюка В. И., Дмитренко М. Е.* Обзор систем управления беспилотных летательных аппаратов общего пользования // Сборник материалов научно-практической конференции «Технологии. Инновации. Связь», Санкт-Петербург, 19 апреля 2021 года. С. 279–286.
2. Стратегия развития отрасли связи Российской Федерации на период до 2035 года [Электронный ресурс] // Правительство России: официальный сайт. 2023. URL: http://government.ru/news/50304/ (дата обращения: 21.05.2024).
3. Юрий Борисов выступил на стратегической сессии, посвящённой серийному производству спутников [Электронный ресурс] // Роскосмос: официальный сайт. 2023. URL: https://www.roscosmos.ru/40096/ (дата обращения: 21.05.2024).
4. Число учтённых в России дронов выросло на треть [Электронный ресурс] // Федеральное агентство воздушного транспорта – Росавиация: официальный сайт, 2024. URL: https://favt.gov.ru/novosti-novosti/?id=10676 (дата обращения: 21.05.2024).
5. *Березкин А. А., Паршин А. А., Лазарев А. А.* Адаптивный контроль интенсивности видеопотока при передаче FPV-трафика беспилотных систем // Материалы 79-й научно-

технической конференции СПб НТО РЭС им. А. С. Попова, посвященной Дню радио, Санкт-Петербург, 22–26 апреля 2024. С. 162–165.
6. *Березкин А. А., Савелов Д. Ю., Суходоева А. В., Туманов И. А., Киричек Р. В.* Исследование нейросетевых моделей предсказания видеопотока при управлении беспилотными системами от первого лица // Труды Научно-исследовательского института радио. 2023. № 3–4. С. 40–56.
7. *Гончаров В. Г., Березкин А. А.* Анализ моделей предсказания кадров FPV-видеопотока в каналах информационного обмена беспилотных систем // Материалы 79-й научно-технической конференции СПб НТО РЭС им. А. С. Попова, посвященной Дню радио, Санкт-Петербург, 22–26 апреля 2024. С. 194–197.
8. *Березкин А. А., Вивчарь Р. М., Киричек Р. В., Захаров А. А.* Метод декомпрессии FPV-видеопотока от беспилотных систем на основе латентной диффузионной нейросетевой модели // Электросвязь. 2024. № 1. С. 42–53.
9. *Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B.* High-resolution image synthesis with latent diffusion models // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022. P. 10684–10695.
10. *Hu X., Huang Z., Huang A., Xu J., Zhou S.* A dynamic multi-scale voxel flow network for video prediction // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023. P. 6121–6131.
11. *Березкин А. А., Вивчарь Р. М., Слеппев А. В., Киричек Р. В., Захаров А. А.* Метод сжатия видеопотока при управлении беспилотными системами в гибридных орбитально-наземных сетях связи // Электросвязь. 2023. № 10. С. 48–56.
12. *Бардасов С. А.* Эконометрика: учебное пособие. 2-е издание, переработанное и дополненное. Тюмень: Издательство Тюменского государственного университета, 2010. 264 с.
13. *Berezkin A. A., Phuc H. Do., Tran D. Le., Kirichek R. V.* A comparative analysis of diffusion-based super-resolution techniques in a video stream compression system in FPV control of unmanned systems // Труды Научно-исследовательского института радио. 2023. № 3–4. P. 26–39.
14. *Козлова А. И., Облаков Н. А., Березкин А. А.* Анализ моделей предсказания кадров FPV-видеопотока в каналах информационного обмена беспилотных систем // Материалы 79-й научно-технической конференции СПб НТО РЭС им. А. С. Попова, посвященной Дню радио, Санкт-Петербург, 22–26 апреля 2024. С. 170–173.

Березкин Александр Александрович
к.т.н., доцент кафедры программной инженерии и вычислительной техники (ПИиВТ), Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича (СПбГУТ, 193232, Санкт-Петербург, пр. Большевиков, д. 22, к. 1), директор Центра перспективных проектов и разработок, e-mail: `berezkin.aa@sut.ru`, ORCID ID: 0000-0002-1748-8642.

Ченский Александр Александрович
магистрант кафедры ПИиВТ, инженер Центра перспективных проектов и разработок, СПбГУТ, e-mail: `chenskii.aa@sut.ru`, ORCID ID: 0009-0005-0832-8590.

Киричек Руслан Валентинович
д.т.н., ректор, профессор кафедры ПИиВТ, СПбГУ, тел. +7 812 3051 200, e-mail: `kirichek@sut.ru`, ORCID ID: 0000-0002-8781-6840.

Авторы прочитали и одобрили окончательный вариант рукописи.
Авторы заявляют об отсутствии конфликта интересов.

Вклад соавторов: *Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.*

Research of Video Stream Intensity Limits in UAV FPV Control in Frame Prediction Mode. Part I: Models and Methods

Alexander A. Berezkin, Alexander A. Chenskiy, Ruslan V. Kirichek

Bonch-Bruevich State university of telecommunications (SPbSUT)

Abstract: Currently, unmanned aerial vehicles have found wide application in various spheres of the national economy. FPV control refers to a method of controlling unmanned aerial vehicles in which the video stream from the unmanned aircraft is transmitted to the remote pilot station in real time. Due to failures in the communication network, packets with video stream data may be lost or delayed and be delivered late. One of the ways to compensate for the drop in FPS in case of frame loss or delay at the remote pilot station is to predict intermediate frames of the video stream. This paper presents a scheme for predicting intermediate frames of a video stream and presents the results of experiments to establish a realistically possible increase in FPS in a neural network codec, in which various autoencoders such as VQ-f16 are used for video compression as well as the lossless latent feature space compression algorithm DEFLATE, while the dynamic multiscale voxel flow neural network DMVFN is used for prediction. A regression model was developed to predict the prediction time. The task of analyzing FPS for various configurations of the neural network decoder on the side of the external pilot is formalized.

Keywords: FPS, variational autoencoder, neural decoder, DMVFN, neural networkы, video stream prediction.

For citation: Berezkin A. A., Chenskiy A. A., Kirichek R. V. Research of video stream intensity limits in UAV FPV control in frame prediction mode. Part I: models and methods (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 3, pp. 115-139. https://doi.org/10.55648/1998-6920-2024-18-3-115-139.

Content is available under the license Creative Commons Attribution 4.0 License

© Berezkin A. A., Chenskiy A. A., Kirichek R. V. 2024

The article was submitted: 05.06.2024;
accepted for publication 20.06.2024.

References

1. Chmelev B. S., Kalyuka V. I., Dmitrenko M. E. Obzor sistem upravlenija bespilotnyh letatel'nyh apparatov obshhego pol'zovanija [Overview of general purpose unmanned aerial vehicle control systems]. *Tehnologii. Innovacii. Svjaz': Sbornik materialov nauchno-prakticheskoj konferencii, Sankt-Peterburg, 19 April, 2021*. Sankt-Peterburg, Federal'noe gosudarstvennoe kazjonnoe voennoe uchrezhdenie vysshego obrazovanija "Voennaja akademija svjazi imeni marshala Sovetskogo Sojuza S. M. Budjonnogo" Ministerstva oborony Rossijskoj Federacii, 2022. pp. 279-286.
2. *Strategija razvitija otrasli svjazi Rossijskoj Federacii na period do 2035 goda* [The strategy for the development of the telecommunications industry of the Russian Federation for the period until 2035], available at: http://government.ru/news/50304/ (accessed: 21.05.2024).
3. *Jurij Borisov vystupil na strategicheskoj sessii, posvjashhjonnoj serijnomu proizvodstvu sputnikov* [Yura Borisov made a presentation at the strategic session on serial production of satellites], available at: https://www.roscosmos.ru/40096/ (accessed: 21.05.2024).
4. *Chislo uchtjonnyh v Rossii dronov vyroslo na tret'* [The number of drones registered in Russia increased by a third], available at: https://favt.gov.ru/novosti-novosti/?id=10676 (accessed: 21.05.2024).
5. Berezkin A. A., Parshin A. A., Lazarev A. A. Adaptivnyj kontrol' intensivnosti videopotoka pri peredache FPV-trafika bespilotnyh sistem [Adaptive control of video stream intensity in FPV traffic transmission of unmanned systems]. *79-ja nauchno-tehnicheskaja konferencija SPb NTO RES im. A.S. Popova, posvjashhennaja Dnju radio: sbornik dokladov regional'noj konferencii*, Saint Peters-

burg,
22-26 April, 2024, pp. 162-165.
6. Berezkin A. A., Savelov D. Ju., Suhodoeva A. V., Tumanov I. A., Kirichek R. V. Issledovanie nejrosetevyh modelej predskazanija videopotoka pri upravlenii bespilotnymi sistemami ot pervogo lica [Research of neural network models for video stream prediction in first-person control in unmanned systems]. *Trudy Nauchno-issledovatel'skogo instituta radio*, 202, no. 3-4, pp. 40-56.
7. Goncharov V. G., Berezkin A. A. Analiz modelej predskazanija kadrov FPV-videopotoka v kanalah informacionnogo obmena bespilotnyh sistem [Analysis of FPV video stream frame prediction models in information exchange links of unmanned systems]. *79-ja nauchno-tehnicheskaja konferencija SPb NTO RES im. A.S. Popova, posvjashhennaja Dnju radio: sbornik dokladov regional'noj konferencii*, Saint-Petersburg, 22-26 April, 2024, pp. 194-197.
8. Berezkin A. A., Vivchar' R. M., Kirichek R. V., Zaharov A. A. Metod dekompressii FPV-videopotoka ot bespilotnyh sistem na osnove latentnoj diffuzionnoj nejrosetevoj modeli [Method of decompression of FPV video stream from unmanned systems based on latent diffusion neural network model]. *Electrosvjaz'*, 2024, №1, pp. 42-53.
9. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10684-10695.
10. Hu X., Huang Z., Huang A., Xu J., Zhou S. A dynamic multi-scale voxel flow network for video prediction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 6121-6131.
11. Berezkin A. A., Vivchar' R. M., Slepnev A. V., Kirichek R. V., Zaharov A. A. Metod szhatija videopotoka pri upravlenii bespilotnymi sistemami v gibridnyh orbital'no-nazemnyh setjah svjazi [Video compression method for controlling unmanned systems in hybrid orbital-terrestrial communication networks]. *Electrosvjaz'*, 2023, no. 10, pp. 48-56.
12. Bardasov S. A. *Ekonometrika: uchebnoe posobie. 2 izdanie, pererabotannoe i dopolnennoe* [Econometrics: textbook. 2nd edition, revised and expanded]. Tjumen', Izdatel'stvo Tjumenskogo gosudarstvennogo universiteta, 2010. 264 p.
13. Berezkin A. A., Phuc H. Do., Tran D. Le., Kirichek R. V. A comparative analysis of diffusion-based super-resolution techniques in a video stream compression system in FPV control of unmanned systems. *Trudy Nauchno-issledovatel'skogo instituta radio*, 2023, no. 3-4, pp. 26-39.
14. Kozlova A. I., Oblakov N. A., Berezkin A. A. Analiz modelej predskazanija kadrov FPV-videopotoka v kanalah informacionnogo obmena bespilotnyh sistem [Analysis of FPV video stream frame prediction models in information exchange channels of unmanned systems]. *79-ja nauchno-tehnicheskaja konferencija SPb NTO RES im. A.S. Popova, posvjashhennaja Dnju radio: sbornik dokladov regional'noj konferencii*, Saint-Petersburg, 22-26 April, 2024, pp. 170-173.

Berezkin Alexander Alexandrovich

PhD, associate professor of the Program Engineering and Computer Science Department, Bonch-Bruevich State university of telecommunications (SPbSUT, 193232, Saint-Petersburg, Bolshevikov Avenue, 22-1), director of the Center of Perspective Projects and Developments, e-mail: berezkin.aa@sut.ru, ORCID ID: 0000-0002-1748-8642.

Chenskiy Alexander Alexandrovich

Master's student of the Program Engineering and Computer Science Department, engineer of Center of Perspective Projects and Developments, SPbSUT, e-mail: chenskii.aa@sut.ru, ORCID ID: 0009-0005-0832-8590.

Kirichek Ruslan Valentinovich

Doctor of technical science, rector, professor of the Program Engineering and Computer Science Department, SPbSUT, phone. +7 812 3051 200, e-mail: kirichek@sut.ru, ORCID ID: 0000-0002-8781-6840.