



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

МИЭМ

МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОНИКИ
И МАТЕМАТИКИ им. А.Н.ТИХОНОВА

МЕЖВУЗОВСКАЯ НАУЧНО-ТЕХНИЧЕСКАЯ КОНФЕРЕНЦИЯ

СТУДЕНТОВ, АСПИРАНТОВ
И МОЛОДЫХ СПЕЦИАЛИСТОВ

им. Е.В.АРМЕНСКОГО



МАТЕРИАЛЫ КОНФЕРЕНЦИИ

2024

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

**МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОНИКИ И МАТЕМАТИКИ им.А.Н.Тихонова
НАЦИОНАЛЬНОГО ИССЛЕДОВАТЕЛЬСКОГО УНИВЕРСИТЕТА
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

**Межвузовская научно-техническая
конференция студентов, аспирантов
и молодых специалистов
имени Е.В. Арменского**

МАТЕРИАЛЫ КОНФЕРЕНЦИИ

Москва 2024 г.

ББК 2+3
Н 34

Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов имени Е.В. Арменского. Материалы конференции. - М. ~: МИЭМ НИУ ВШЭ, 2024. – 418 стр.

ISBN 978-5-903650-78-1

В материалах конференции студентов, аспирантов и молодых специалистов представлены тезисы докладов по ключевым научным направлениям Московского института электроники и математики им. А.Н.Тихонова: телекоммуникационным технологиям и системам связи, методам и системам защиты информации, исследованию электронных компонентов систем и технологий, математическому и компьютерному моделированию классических и квантовых систем. Тезисы докладов разделены по следующим тематическим секциям, исследованию классических и квантовых материалов:

- Аналитика, статистика и искусственный интеллект
- Компьютерное моделирование физико-механических систем
- Математическое моделирование и динамические системы
- Компьютерные сети, телекоммуникационные системы и интернет вещей
- Материалы и технологии
- Низкоразмерные квантовые системы
- Функциональные квантовые материалы
- Кибербезопасность
- Микроволновая и оптическая электроника, электромагнитная совместимость
- Электронное приборостроение

Материалы конференции могут быть интересны для преподавателей, студентов, научных сотрудников и специалистов в области прикладной математики, информационно-коммуникационных технологий, электроники и информационной безопасности.

Редакционная коллегия: Е.А. Крук, С.А. Аксенов, К.Ю. Арутюнов, Г.Г. Бондаренко, Л.С. Восков, В.Г. Данилин, О.О. Евсютин, А.А. Елизаров, С.А. Слестников, С.И. Нефедов, Н.С. Титкова, А.А. Шаненко

Издание осуществлено с авторских оригиналов.

ISBN 978-5-903650-78-1

© Московский институт электроники и математики Национального исследовательского университета «Высшая школа экономики», 2024 г.
© Авторы, 2024 г.

Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов им. Е.В. Армения названа в честь основателя и первого ректора Московского института электроники и математики. Евгений Викторович в 1962 году создал вуз на стыке самых передовых образовательных методик того времени в области электроники и реального сектора экономики, представленного крупнейшими предприятиями в этой области.

Краткая биографическая справка: Евгений Викторович Армения родился 2 октября 1923 года в городе Буй Костромской области. В 1941 году он ушел добровольцем на фронт в составе 234-й Ярославской стрелковой коммунистической дивизии. Первый фронтовой опыт получил в боях под Москвой. В самые тяжелые 1941-1942 годы он был разведчиком-наблюдателем 1081 артполка на Калининском фронте, а после ранения в конце 1942 года - курсантом учебной роты связи 5-ой отдельной запасной стрелковой бригады Уральского военного округа. С 1943 года по ноябрь 1945 года был командиром радиовзвода 102 стрелкового корпуса Первого Украинского фронта. Великую Отечественную войну Евгений Викторович закончил в Праге 9 мая 1945 года. За ратные подвиги он был награжден орденами Красной Звезды, Отечественной войны I и II степени, медалями «За боевые заслуги», «За освобождение Праги», «За победу над Германией в Великой Отечественной войне».

После демобилизации он поступил в Ивановский химико-технологический институт, в котором проучился 2 курса, а затем перешел на третий курс Московского механического института (ныне Московский инженерно-физический институт, НИЯУ МИФИ), который закончил в 1951 году.

Трудовую деятельность Евгений Викторович начал с должности секретаря комитета ВЛКСМ МИФИ, а с 1952 года началась его педагогическая деятельность ассистентом кафедры «Автоматика и телемеханика» МИФИ. В этом же году он поступил в аспирантуру МИФИ, которую успешно завершил, и защитил кандидатскую диссертацию на тему «Пространственный метод регулирования двухфазного асинхронного двигателя». С 1956 года Евгений Викторович развивает научное направление и формирует научную школу в области создания систем автоматического контроля, регулирования и управления электрофизических установок.

12 апреля 1962 года Советское правительство приняло решение о создании Московского института электронного машиностроения. Евгений Викторович Армения был назначен ректором созданного института. С этого времени в полной мере раскрылись выдающиеся организаторские и педагогические способности Евгения Викторовича. В МИЭМ, благодаря его кипучей и самоотверженной деятельности, были организованы новые кафедры и научные лаборатории, сформирован дружный коллектив единомышленников. В начале 70-ых годов прошлого столетия в МИЭМ по инициативе академика АН СССР С.Н. Вернова и профессора Е.В. Армения впервые в нашей стране была организована подготовка специалистов в области космического и радиационного материаловедения, созданы лаборатории электронной микроскопии, рентгеновского анализа.

Евгений Викторович Армения уделял большое внимание учебной и научной деятельности. В 1968 году выходит в свет учебное пособие Е. В. Армения и Г.Б. Фалка



«Электрические микромашины». В 1969 году Евгений Викторович защищает докторскую диссертацию на тему «Системы управления ускорителями заряженных частиц с энергией до 100 МэВ», в этом же году ему присвоено ученое звание «профессор» по кафедре «Электрические машины и электромеханические устройства автоматики». Профессор Евгений Викторович Арменский хорошо известен своими монографиями и учебниками, изданными на русском, английском, французском, испанском языках. Эти книги стали классическими настольными пособиями для специалистов и студентов практически во всех научных и учебных центрах.

В 1973 году Евгению Викторовичу присвоено почетное звание «Заслуженный деятель науки и техники РСФСР». В 1984 году ему вместе с коллегами была присуждена Государственная премия СССР «За создание и внедрение в промышленность сверхвысоковакуумных магниторазрядных насосов и высоковакуумных средств технологического и научного оборудования электронной техники».

Евгений Викторович всегда уделял большое внимание подготовке молодых ученых. Под его руководством выполнены и защищены 10 докторских и свыше 30 кандидатских диссертаций. На протяжении 40 лет Евгений Викторович был членом редколлегии журналов «Измерительная техника», «Физика и химия обработки материалов» и «Датчики и системы».

В своей педагогической деятельности Евгений Викторович Арменский всегда призывал студентов, аспирантов стремиться получить дополнительные знания в области прикладной и фундаментальной математики, а также использовать современные информационные технологии в научной работе.

Секция «Аналитика, статистика и искусственный интеллект»

АНАЛИЗ АЛГОРИТМОВ ОБНАРУЖЕНИЯ ДИПФЕЙКОВ

Федотов Г.А.

*Национальный исследовательский университет
«Высшая школа экономики»,
аспирантская школа по техническим наукам*

Аннотация

В последние годы наблюдается значительный прогресс в качестве синтетически сгенерированного контента. Регулярно появляются инструменты, с помощью которых обычный пользователь ПК может создать реалистичный поддельный контент. В работе исследуется развитие генеративных моделей в задаче Face Synthesis, а также способы обнаружения дипфейков, созданных с помощью моделей этого класса.

Введение

В современном мире с огромным количеством интернет-ресурсов очень легко распространять ложную информацию [6]. Для ее генерации активно используются методы, которые позволяют синтезировать реалистичный мультимедийный контент – дипфейки. Многие решения по генерации дипфейков получили удобный пользовательский интерфейс (Zao[1], REFACE[2], FaceApp[3], Audacity [4], Soundforge [5]), что позволило обычным пользователям использовать их в своих целях. Естественно, этократно увеличило количество дипфейков в медиапространстве.

Возможные области применения дипфейков: шантаж людей с целью финансовой выгоды, создание политических или религиозных волнений. Ожидается, что дипфейки активно будут использоваться следующие социальные группы:

«Интернет-тролли» - пользователи Интернета, которые распространяют подстрекательскую информацию, пытаясь вызвать реакцию конкретного человека или группы людей.

Желтая пресса - веб-сайты и блоги с фейковыми новостями, слухами и сплетнями, которые намеренно распространяют ложную информацию.

Группы людей с политическими интересами - публикация политической пропаганды против стран, партий или конкретных лиц. Это могут быть: фейковые обращения политика к народу, фейковые «слитые» в сеть разговоры публичных лиц, а также более сложные сценарии.

Мошенники с помощью дипфейков смогут выдавать себя за знакомых жертвы или публичных людей, «попавших» в сложную ситуацию или «собирающих» средства для благих целей.

Таким образом, использование дипфейков может нанести серьезный ущерб, а их обнаружение является актуальной и важной задачей.

Обзор алгоритмов в задаче Face Synthesis

Задача Face Synthesis предполагает синтез фотореалистичных изображений человеческого лица, которые могут существовать, а могут и не существовать в реальной жизни. Классическими методами генерации изображения с помощью методов глубокого обучения можно назвать архитектуры GAN [7] и VAE [8]. Изображения, созданные с помощью этих архитектур, были низкого разрешения и содержали заметные искажения - «артефакты». Однако, именно они сделали первый прорыв в этой области и их архитектурные идеи до сих пор служат основой для современных генеративных моделей.

Вариационный автоэнкодер (VAE) [8] основывается на идее автоэнкодера (AE) - нейросети, которая состоит из двух частей - энкодера и декодера. Энкодер сжимает изображение в низкую размерность, называемую скрытым представлением, а декодер разжимает его обратно и пытается получить такую же картинку, какая была на входе у энкодера. При успешном обучении такой сети ожидается, что скрытое представление будет довольно хорошо описывать признаки входного объекта и разумно отбрасывать ненужную информацию - например, шум. Классические автоэнкодеры обучаются кодировать входные данные и восстанавливать их, однако получившееся скрытое пространство может не быть непрерывным и не позволять производить интерполяцию. Вариационный автоэнкодер (VAE) [8] имеет одно уникальное свойство, которое отличает его от стандартного автоэнкодера. Именно это свойство делает вариационные автоэнкодеры столь полезными при генерации данных: их скрытое пространство по построению является непрерывным, позволяя выполнять случайные преобразования и интерполяцию. Непрерывность скрытого пространства достигается следующим образом: для каждого входного X энкодер предсказывает два значения - среднее μ и стандартное отклонение σ , с помощью которых строится нормальное распределение. Для генерации изображения нам достаточно сгенерировать z из, полученного на прошлом шаге, нормального распределения и подать его на вход декодеру.

Генеративно-состязательная сеть (GAN) [7] состоит из двух нейронных сетей - генератора G и дискриминатора D . Задача генератора состоит в генерации новых образцов. Дискриминатор представляет из себя бинарный классификатор, задача которого оценить вероятность, что полученный на вход образец пришел из тренировочного набора данных, а не был создан генератором G . Обучение модели G состоит в максимизации вероятности ошибки дискриминатора D . Другими словами, в процессе обучения G и D играют в «минмакс» игру.

С момента появления GAN [7] в 2014 г. были предприняты значительные усилия по улучшению

качества синтезируемых изображений. Так, DCGAN[8] был первым подходом, который ввел в генератор слой деконволюции вместо полносвязного слоя, что улучшило его производительность.

Karras T. и др. [9] представили ProGAN, на каждом этапе обучения которого нейросеть обучалась создавать картинки вчетверо больше, начиная от 4×4 . Как правило, конечный результат имел размерность 1024×1024 . Данный подход позволил существенно сэкономить время для обучения модели.

StyleGAN [10] — это улучшенная версия ProGAN [9]. Новая архитектура разделяет атрибуты высокого уровня (напр., позы) и стохастическим изменениям в сгенерированных изображениях (напр., веснушки, волосы) StyleGAN [10] достиг высокого разрешения сгенерированных изображений, т.е. 1024×1024 , с проработанными мелкими деталями.

StyleGAN2 [11] улучшил реалистичность изображения за счет удаления нежелательных артефактов, таких как корректировка направления взгляда, положение зубов и выражение лица.

Способы детекции Face Synthesis

Задача детекции дипфейков фактически является задачей бинарной классификации. Таким образом, метриками для оценки качества модели используются стандартные метрики для оценки качества бинарного классификатора: Accuracy, AUC, ROC, Error rate, Precision, Recall, Sensibility, TPR, F1 score. Для вычисления этих метрик необходимо построить матрицу ошибок, в которой TP (true positive) - количество дипфейков, которые были правильно предсказаны, а TN (true negative) - количество реальных изображений/видео, правильно предсказанных как настоящие. FP (false positive) - количество реальных изображений/видео, которые модель некорректно предсказала, как Deepfake, FN (false negative) - количество Deepfake, неправильно предсказанных как реальные изображения/видео.

Детекция дипфейков, сгенерированных с помощью методов глубокого обучения, является новой задачей. Поэтому в первых экспериментах в этой области старались использовать опыт выявления поддельных изображений [12, 13]. Однако, эти методы недостаточно хорошо справлялись с поставленной задачей.

Guarnera L. и др. предложили [14] извлекать признаки изображения с помощью EM алгоритма, а затем передавать их на вход классификатору (авторы статьи использовали KNN, SVM, LDA). Данный подход с точностью >99% обнаруживал дипфейки, созданные с помощью StyleGAN и StyleGAN2.

В статье [15] представлен способ применения биспектрального анализа в задаче по обнаружению дипфейков, сгенерированных StyleGAN2. Эксперимент проводился на двух наборах данных. Результаты показали, что спектральные корреляции более высокого порядка, выявленные с помощью спектрального анализа, меньше присутствуют в реальных изображениях, чем в дипфейках. Данный подход с

точностью ~80% обнаруживал дипфейки, созданные с помощью StyleGAN2.

В своей работе Nataraj L. и др. предложили [16] вычислять матрицы совместного появления изображения по трем цветовым каналам и использовать их для тренировки CNN. Подобный подход показал многообещающие результаты для распознавания дипфейков, созданных некоторыми нейронными сетями на основе GAN. Данный подход с точностью 97.84% обнаруживал дипфейки, созданные с помощью CycleGAN и StarGAN.

Таблица 1. Сравнение методов обнаружения дипфейков

| Признаки | Генеративная модель | Точность определения |
|--|---------------------|----------------------|
| Признаки изображения, полученные с помощью EM алгоритма, подаются на вход SVM | StyleGAN, StyleGAN2 | 99% |
| Биспектральный анализ | StyleGAN2 | 80% |
| Вычисление матрицы совместного появления и использование ее для тренировки CNN | CycleGAN, StarGAN | 97.84% |

Для задачи обнаружения дипфейков существуют следующие наборы данных [17]: FaceForensics, FaceForensics++, Deepfake Detection, DeepFake Forensics V1, DeepFake Forensics V2, Deepfake Detection Challenge, Deepfake-TIMIT, DeeperForensics-1.0, Wild Deepfake, SwapMe and FaceSwap, Deep Fakes, Fake Faces in the Wild, FakeET, Face Shifter, Deepfake, Swapped Face Detection, Inconsistent Head Poses, Tampered Face.

Заключение

В данной работе представлен ряд методов, с помощью которых можно обнаруживать дипфейки. Каждый метод обладает своими преимуществами и недостатками, поэтому для создания качественного решения рекомендуется комбинировать различные подходы. Крайне быстрые темпы развития генеративных алгоритмов способствуют созданию все более качественных дипфейков, для которых необходимо заново тестировать описанные выше техники обнаружения. Дальнейшие исследования предполагается проводить в следующих направлениях: поиск новых техник обнаружения синтетически сгенерированного мультимедийного контента, рассмотрение других типов дипфейков, таких как Face-swap, Lip-syncing, Puppet-master, Facial Attribute Manipulation, Audio Deepfakes, проверка применимости старых техник обнаружения дипфейков к мультимедийному контенту, созданному с помощью новейших генеративных алгоритмов.

Список литературы

1. ZAO : официальная страница в App Store. — Текст : электронный // App Store : [сайт]. — URL: <https://apps.apple.com/cn/app/zao/id1465199127> (дата обращения: 01.02.2024).
2. Reface App : официальный сайт. — Текст : электронный // Reface App : [сайт]. — URL: <https://reface.app/> (дата обращения: 01.02.2024).
3. FaceApp : официальный сайт. — Текст : электронный // FaceApp : [сайт]. — URL: <https://www.faceapp.com/> (дата обращения: 01.02.2024).
4. Audacity : официальный сайт. — Текст : электронный // Audacity : [сайт]. — URL: <https://www.audacityteam.org/> (дата обращения: 02.02.2024).
5. Sound Forge : официальная страница. — Текст : электронный // Magix : [сайт]. — URL: <https://www.magix.com/gb/music/sound-forge/> (дата обращения: 02.02.2024).
6. Figueira A., Oliveira L. The current state of fake news: challenges and opportunities // *Procedia computer science*. – 2017. – Т. 121. – С. 817-825.
7. Creswell A. et al. Generative adversarial networks: An overview // *IEEE signal processing magazine*. – 2018. – Т. 35. – №. 1. – С. 53-65.
8. Radford A., Metz L., Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks // *arXiv preprint arXiv:1511.06434*. – 2015.
9. Karras T. et al. Progressive growing of gans for improved quality, stability, and variation // *arXiv preprint arXiv:1710.10196*. – 2017.
10. Karras T. et al. Analyzing and improving the image quality of stylegan // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. – 2020. – С. 8110-8119.
11. Huang R. et al. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis // *Proceedings of the IEEE international conference on computer vision*. – 2017. – С. 2439-2448.
12. Zhang Y., Zheng L., Thing V. L. L. Automated face swapping and its detection // *2017 IEEE 2nd international conference on signal and image processing (ICSIP)*. – IEEE, 2017. – С. 15-19.
13. Akhtar Z., Dasgupta D. A comparative evaluation of local feature descriptors for deepfakes detection // *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*. – IEEE, 2019. – С. 1-5.
14. Guarnera L., Giudice O., Battiato S. Deepfake detection by analyzing convolutional traces // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. – 2020. – С. 666-667.
15. С. П. Никитенкова Применение биспектрального анализа в обнаружении deepfake-изображений // *Системы и средства информ.* – 2023). – С. 25–33
16. Nataraj L. et al. Detecting GAN generated fake images using co-occurrence matrices // *arXiv preprint arXiv:1903.06836*. – 2019.
17. DeepFake Detection Datasets. — Текст : электронный // paperswithcode : [сайт]. — URL: <https://paperswithcode.com/datasets?task=deepfake-detection> (дата обращения: 08.02.2024).

РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ ЭМОЦИЙ

Коровин Е.А., Чиглинцева С.А.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В исследовательской работе представлен алгоритм распознавания эмоций. Был проведен обзор существующих методов и средств для решения подобных задач. Для достижения поставленной цели использовались инструменты библиотеки OpenCV в языке программирования Python 3. Результаты экспериментов подтверждают конкурентоспособность разработанной модели в сравнении с аналогичными решениями, а точность распознавания эмоций превышает 68%.

Введение

Прогресс в области информационных технологий, внедрение которых в повседневную жизнь обычных граждан продолжает наращаться, оказывает значительное воздействие не только на внешние аспекты их жизнедеятельности, подразумеваемые термином "цифровизация", но также на их поведенческие характеристики. Этот вопрос стимулирует интерес к тому, как с использованием текущего технологического прогресса, обусловленного накоплением обширных данных и увеличением доступности индивидуальных средств взаимодействия с цифровой средой, возможно обрести новые методы понимания, взаимодействия и воздействия как на группы людей, так и на индивида.

В рамках таких исследований анализ данных, включая распознавание эмоций по изображению лица человека, приобретает все большую актуальность. Это обусловлено увеличивающимся интересом к пониманию человеческих эмоций как в сфере маркетинга и рекламы, так и в психологии, медицине и других дисциплинах.

Исследования в области распознавания эмоций обладают значительным потенциалом для практического применения в различных областях. Например, в медицине они могут быть полезны для диагностики психических расстройств, анализа эмоционального состояния пациентов и мониторинга их прогресса. В области маркетинга и рекламы они способны улучшить понимание реакций потребителей на продукты и услуги, помогая компаниям адаптировать свои стратегии к их потребностям и, таким образом, оптимизировать коммуникацию между ними.

Описательная постановка цели и задачи

Целью данной работы является разработка системы распознавания эмоций по фотографии. В ходе анализа поставленной задачи, была выявлена необходимость создания полноценного многоэтапного алгоритма, который включал бы в себя не только определение эмоции по фото лица человека, но и

этапы поиска на изображении лиц, выделение каждого лица в отдельности и предоставление результатов распознавания эмоции по каждому из найденных лиц. В качестве требований к разрабатываемому решению выдвигается: показатель точности распознавания, который должен являться сопоставимым или превышающим существующие публичные решения, а время распознавания, что не должно превышать одной секунды.

Процесс распознавания эмоций представляет собой последовательный набор этапов. На первом этапе осуществляется локализация лица на изображении, где определяется точное положение лица среди других объектов на фотографии или видеозаписи.

Далее следует этап нормализации изображения лица, где происходит стандартизация изображения с целью устранения возможных искажений, таких как повороты или изменения освещения, что позволяет достичь более точного и унифицированного представления лица для последующего анализа.

Последний этап – классификация, где производится определение к какому из определенных классов эмоций относится локализованное и прошедшее ранее необходимую предобработку изображение лица.

Рассмотрим подробнее процесс, представленный на рис. 1.

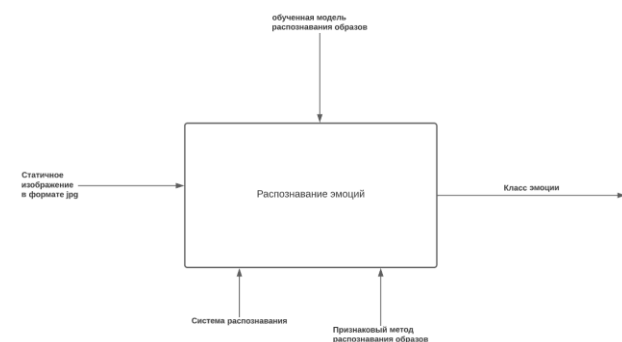


Рис.1. Функциональная модель IDEF (уровень 0)

На рис.2 представлена IDEF 1 уровня декомпозиции.



Рис.2. Диаграмма декомпозиций IDEF (уровень 1)

Обзор существующих методов распознавания эмоций

В процессе методологии принятия решений при распознавании образов выделяются три основных направления: *эвристические методы*, *математиче-*

ские методы и *лингвистические (синтаксические) методы*.

Эвристические методы, такие как перцептрон, базируются на опыте и интуиции разработчика системы. Их основным недостатком является отсутствие гарантии успешного решения задачи, что может привести к неопределенности при неудачных попытках применить интуитивные соображения.

Математические методы опираются на использование классического математического аппарата, включая методы линейного программирования, корреляционного анализа, и т.д. Среди них выделяются методы решающих функций, статистический метод, метод расстояния Хаусдорфа, метод взаимного отклонения и метод расстояния Хэмминга.

Лингвистические (синтаксические, структурные) методы применяются в случаях, когда образ представляет собой структуру из неприводимых элементов и признаков, описывающих связи между ними. Здесь широко используется аппарат алгебры логики и теории формальных языков.

В контексте методов распознавания эмоций рассмотрим классический способ определения эмоции по изображению лица, основанный на классификации ключевых точек, координаты которых можно получить, используя различные алгоритмы PDM, CML, AAM и другие. На изображении лица размечается множество точек, привязанных к положению бровей, глаз, губ, носа и челюсти, что позволяет осуществить частичный захват мимики. Нормализованные координаты точек подаются в классификатор (например, Random Forest) и получается базовое решение. Естественно положение лиц при этом должно быть выровнено. Однако даже при выполнении всех условий достигаемая точность будет отставать от более современных решений, поэтому данный метод можно назвать в определенном смысле устаревшим.

Также рассмотрим другой метод, основанный на геометрических характеристиках лица — это метод, предложенный Полом Экманом в их классической работе "Распознавание лицевых экспрессий" (позже FACS). Этот метод использует множество ключевых точек на лице, которые соответствуют различным мимическим мышцам. На основе положения этих точек и вычисления различных геометрических параметров, таких как расстояния, углы и площади, алгоритм может классифицировать основные эмоциональные состояния, которых, согласно идеологу данного подхода существует семь. Одной из основных проблем данного метода является изначальная его ориентированность на применение человеком, от того компьютерный расчет двигательных единиц может быть весьма затруднителен. Так, наличие препятствий в виде макияжа, татуировок, необычной прически и других может помешать поиску необходимых точек, а, следовательно, и процессу распознавания.

Еще одним методом является метод главных компонент, который представляет изображения лиц в виде набора главных компонент. В задачах распознавания эмоций этот метод может использоваться для уменьшения размерности данных, без значительной

потери информации об эмоциональных характеристиках, что позволяет сократить вычислительные затраты, тем самым влияя на быстрдействие решения в целом.

Сверточная нейронная сеть использует особенности зрительной коры и обладает высокой точностью распознавания, которая может достигать значений более 90%. В числе прочих плюсов данного подхода стоит отметить отсутствие человеческого фактора при извлечении признаков., что отличает его от более традиционных подходов, где зачастую применяется ручная разработка признаков. Что важно, свёрточные нейронные сети обладают способностями к моделированию пространственных взаимосвязей и извлечению глубинных зависимостей, что дает методу возможность работы с глубинными знаниями и особенностями изображений, за счёт чего его результаты зачастую оказываются выше тех, что могут быть получены другими методами. Однако ее реализация затруднена, и существует риск переобучения при добавлении новых эталонных данных в базу.

Поиск исходных данных для разработки решения

Существуют несколько широко используемых наборов данных, содержащих изображения лиц людей с различными эмоциями. Один из наиболее известных - "FER2013" (The Facial Expression Recognition 2013), включает более 35 тысяч изображений лиц размером 48 на 48 пикселей в оттенках серого, разделенных на семь классов эмоций: радость, грусть, удивление, страх, отвращение, гнев и нейтральное состояние. Подготовка данных производилась таким образом чтобы лицо располагалось более или менее по центру и занимало примерно одинаковое, превышающее 50%, пространство на каждом изображении. Этот набор данных широко применяется для обучения и тестирования моделей распознавания эмоций.

Другой популярный набор данных "СК+" (Extended Cohn-Kanade) включает изображения лиц с различными эмоциями и информацию о движении мышц лица, способствуя более глубокому пониманию взаимосвязи между эмоциями и физиологическими процессами в лице.

Тем не менее, несмотря на наличие этих наборов данных, остаются незакрытые потребности в повышении качества обучающих наборов для распознавания эмоций. Эти потребности включают в себя несбалансированность классов эмоций, разнообразие культурных контекстов и выражений лиц, а также необходимость учета динамики эмоций во времени.

С развитием методов и подходов, использование упомянутых наборов данных и их развитие становится все более распространенной задачей в научных и практических исследованиях.

Описание разработанного решения

Благодаря проведенному анализу был выбран подход, основанный на использовании сверточной нейронной сети, для разработки собственного решения, поскольку данный нейросетевой метод эффек-

тивен в решении слабоструктурированных задач. Основным преимуществом выбранного метода является способность моделировать сложные нелинейные зависимости между входными и выходными переменными без явного установления точной формульной зависимости.

В качестве датасета для обучения предлагаемой модели был выбран "FER2013" (The Facial Expression Recognition 2013), несмотря на высокую величину размера датасета для обучения сверточных нейронных сетей (приблизительно 35 тысяч изображений) и описанную несбалансированность классов эмоций. На данный момент этот датасет представляет собой наиболее конкурентоспособное решение по сравнению с аналогичными доступными наборами данных.

Архитектура сети представлена на рис.3 Разработка собственной сети позволила решить определенные проблемы, в частности, обеспечить наличие лица на изображении при тестировании во всех случаях. Это означает, что при непосредственных испытаниях программы не возникало ситуаций, когда изображение с лицом человека расценивалось как изображение без лица, и была предоставлена возможность измерения характеристик сети.

Разработанное решение обладает точностью распознавания, превышающей 68%, что сопоставимо с другими публично доступными решениями для данного датасета и временем распознавания не превышающем заданный порог в 1 секунду.

```
Model: "DCNN"
```

| Layer (type) | Output Shape | Param # |
|-----------------------------------|---------------------|---------|
| conv2d_1 (Conv2D) | (None, 48, 48, 64) | 1664 |
| batchnorm_1 (Batch Normalization) | (None, 48, 48, 64) | 256 |
| conv2d_2 (Conv2D) | (None, 48, 48, 64) | 182464 |
| batchnorm_2 (Batch Normalization) | (None, 48, 48, 64) | 256 |
| maxpool2d_1 (Max Pooling 2D) | (None, 24, 24, 64) | 0 |
| dropout_1 (Dropout) | (None, 24, 24, 64) | 0 |
| conv2d_3 (Conv2D) | (None, 24, 24, 128) | 73856 |
| batchnorm_3 (Batch Normalization) | (None, 24, 24, 128) | 512 |
| conv2d_4 (Conv2D) | (None, 24, 24, 128) | 147584 |
| batchnorm_4 (Batch Normalization) | (None, 24, 24, 128) | 512 |
| maxpool2d_2 (Max Pooling 2D) | (None, 12, 12, 128) | 0 |
| dropout_2 (Dropout) | (None, 12, 12, 128) | 0 |
| conv2d_5 (Conv2D) | (None, 12, 12, 256) | 295168 |
| batchnorm_5 (Batch Normalization) | (None, 12, 12, 256) | 1024 |
| conv2d_6 (Conv2D) | (None, 12, 12, 256) | 590880 |
| batchnorm_6 (Batch Normalization) | (None, 12, 12, 256) | 1024 |
| maxpool2d_3 (Max Pooling 2D) | (None, 6, 6, 256) | 0 |
| dropout_3 (Dropout) | (None, 6, 6, 256) | 0 |
| flatten (Flatten) | (None, 9216) | 0 |
| dense_1 (Dense) | (None, 128) | 1179776 |
| batchnorm_7 (Batch Normalization) | (None, 128) | 512 |
| dropout_4 (Dropout) | (None, 128) | 0 |
| out_layer (Dense) | (None, 7) | 983 |

Рис.3. Архитектура собственной сети

Сравнительная оценка решения относительно существующих разработок

Полученные результаты требуется оценить, для этого прибегнем к методу, который предполагает сравнение с сопоставимыми аналогами. В качестве объекта сравнения выбрана исследовательская работа с аналогичной тематикой, однако с отличающимся методом разработки нейронной сети [4]. В данном случае, для проведения сравнительного анализа использованы результаты исследования, где применялась нейронная сеть с переносом обучения. Это подразумевает использование предварительно обученной модели ResNet50v2 (приблизительно 50 миллионов изображений) с измененным выходным классификатором, обученным на используемом в нашем исследовании датасете.

В рамках данного исследования, после прохождения 112 эпох точность работы модели достигла 73%. Результат данного эксперимента, хотя и высокий, сопровождается вероятностью переобучения модели. Под переобучением понимается способность модели, особенно при малом количестве обучающих данных, «запоминать» их, что позволяет ей показывать лучшие результаты на этих данных, однако это приводит к ухудшению способностей модели к обобщению, из-за чего реальная предсказательная сила модели может оказаться ниже полученной при обучении точности. Так как в описанном исследовании не проведена дополнительная проверка результатов на неиспользованных в процессе обучения изображениях, для частоты эксперимента ориентируемся на результаты, полученные моделью к моменту 35–40 эпох, с соответствующим уровнем точности в пределах 67-69%. Данный промежуток был выбран так как на нем начинается видимое расхождение между обучающей и валидационной ошибкой, где обучающая ошибка продолжает сокращаться, а валидационная выходит на «плато» с разнонаправленными колебаниями показателя, что указывает на возникновение на этом этапе переобучения. Взятый для оценки результат, соответствующий описанному эпохальному периоду - согласуется с нашими результатами.

Также было проведено сравнение с наиболее популярными решениями, представленными на форуме Kaggle. Показатели, достигнутые в этих решениях, либо ниже, либо условно сопоставимы с достигнутыми нашей моделью показателями. Это позволяет говорить о конкурентоспособности разработанной модели.

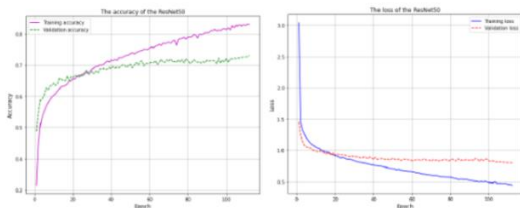


Рис. 4. Изменение метрик «accuracy» и «loss» на обучающем и валидационном наборах данных в процессе обучения нейронной сети, построенной на основе ResNet50

Алгоритм полного распознавания, начиная от изображения и завершая выявлением эмоций, разработан с использованием инструментария библиотеки OpenCV для языка программирования Python 3. Средствами данной библиотеки осуществляется поиск лиц на входящих фотографиях и их предварительная обработка с соответствием стандартам нашей модели. Эта обработка включает в себя выделение квадратной области, содержащей лицо, перевод изображения в черно-белую гамму и изменение размера до 48 на 48 пикселей, что соответствует входным параметрам нашей модели. Графическое представление полного алгоритма распознавания изображено на рис.5.

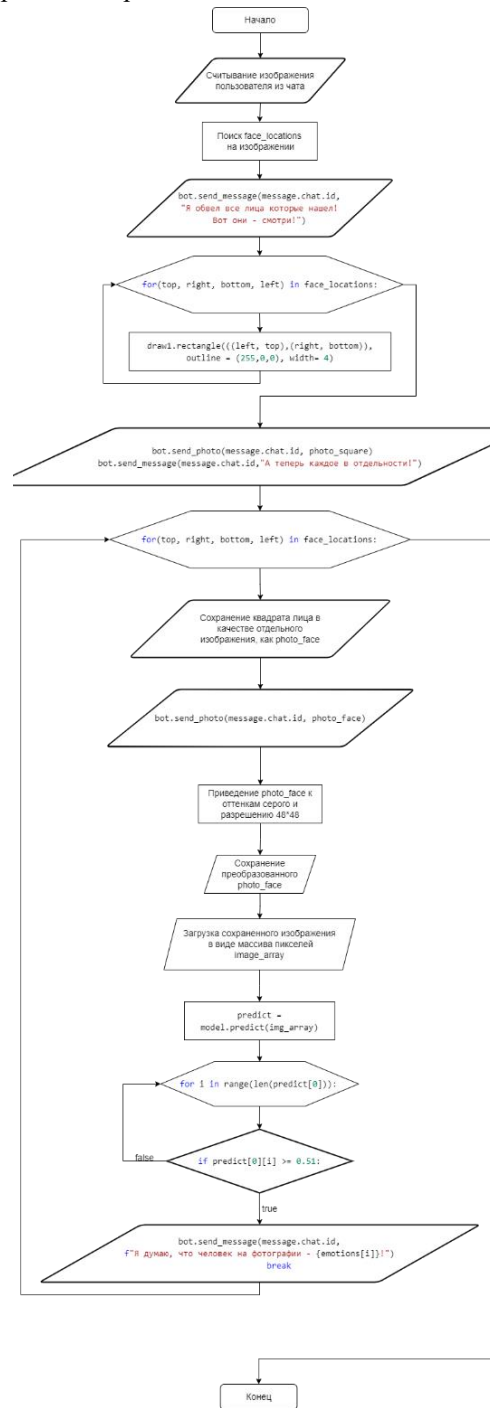


Рис. 5. Блок-схема работы алгоритма распознавания

Заключение

В рамках проведенного исследования были рассмотрены существующие подходы к решению задачи распознавания эмоций и проанализированы датасеты, подходящие для этого. Было разработано программное решение, воплощающее в себе алгоритм распознавания полного цикла, от поступления на вход неподготовленного изображения, до, в качестве выхода, определения класса эмоции для каждого представленного на изображении лица. В своей специфике оно основано на применении сверточной нейронной сети собственной архитектуры. Исследование имеет перспективы для применения в различных сферах, от использования в качестве научной базы будущих работ в области, до, непосредственно применения в процессах диагностики психических расстройств в медицине и улучшения стратегий в маркетинге и рекламе.

Список литературы

1. Viola, P. Rapid object detection using a boosted cascade of simple features // IEEE Conf. on Computer Vision and Pattern Recognition. –Kauai, Hawaii, 2008. – V. 1. – p. 513–520.
2. Viola, P. Robust real-time face detection // International Journal of Computer Vision. – 2005. –V. 58. – № 2. – p. 137–155.
3. Метод Виолы-Джонса (Viola-Jones) как основа для распознавания лиц. URL: <https://habrahabr.ru/post/133826/> (Дата обращения 02.02.2024)
4. Галичий Д.А., Афанасьев Г.И., Нестеров Ю.Г. Распознавание эмоций человека при помощи современных методов глубокого обучения // E-Scio. 2021. №5 (56). URL: <https://cyberleninka.ru/article/n/raspoznavanie-emotsiy-cheloveka-pri-pomoschi-sovremennyh-metodov-glubokogo-obucheniya> (дата обращения: 30.01)

СПОСОБНОСТЬ РАСПОЗНАВАНИЯ СИСТЕМЫ ГОЛОСОВОГО ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ В УСЛОВИЯХ ВОЗДЕЙСТВИЯ ШУМОВ И ПОМЕХ

Тарадайко Е.А., Тузов Н.А.

Севастопольский государственный университет, кафедра «Информационная безопасность»

Аннотация

В данной статье приведены краткие результаты исследований возможностей распознавания речи человека отечественными системами голосового человеко-машинного взаимодействия (СГЧМВ) в условиях разного рода помех. Данные результаты могут быть использованы для улучшения их способностей распознавания, а также определения особенностей данного процесса, в том числе результаты будут полезны для сопоставления результатов распознавания

речи человека и СГЧМВ в идентичных условиях зашумления.

Введение

Важность исследования возможностей распознавания речи человека отечественными СГЧМВ обусловлена их быстрым развитием и внедрением практически во все сферы жизни человека. Общеизвестным фактом является, что данные системы установлены на каждом современном смартфоне и компьютере, а у многих СГЧМВ даже управляют домом.

Развитие СГЧМВ, непосредственно связано с акустической средой. При этом СГЧМВ, приходится воспринимать как модель черного ящика со множеством неизвестных параметров. Естественно, для анализа корректности использования СГЧМВ, в большинстве случаев применяют моделирование и в редких случаях натуральный эксперимент. В большинстве случаев при моделировании используют синтезированную помеху вместо звукозаписи естественных шумов акустической среды.

В работе изучается способность распознавания русской речи СГЧМВ при добавлении к звукозаписи естественного шума (взят частный случай шума «звукозапись разговора десяти человек в помещении»), и случаи добавления синтезированной помехи (созданной в среде матлаб белый шум с равномерным законом распределения), в зависимости от их интенсивности.

В данной работе все результаты представлены в единицах измерения дБ. Для их перевода из единицы измерения «фазы» и упрощения понимания читателем откуда в дальнейшем появляется знак «-», необходимо воспользоваться формулой 1.

$$1\text{дБ} = 20 \log_{10} \left(\frac{\|x\|}{\|u\|} \right), \quad (1)$$

где x – мощность сигнала в разгах; u – мощность шума в разгах.

Исследования

Редко встречающимся в литературе методом оценки распознавания речи СГЧМВ, является использование инструмента словесной разборчивости [1, 2, 4], как наиболее приближенного к восприятию человеком. Так как СГЧМВ способны не только распознавать текст из речи, но дополнять при транскрипции и исправлять ошибки. Отметим, что в описанных работах, отсутствует сравнение оценки работы СГЧМВ в условиях естественных шумов.

Модель, на основе которой проводятся эксперименты, состоит из двух основных блоков: первый блок – создание специализированного речевого корпуса, а второй – непосредственное проведение экспериментов, которое заключается в циклическом прогоне через СГЧМВ зашумленных записей с разным коэффициентом сигнал/шум [5].

В эксперименте использовался русскоязычный речевой корпус, состоящий из простых фраз, выбранных из ГОСТ 16600-72 [1], что позволило так же соизмерить способность распознавания СГЧМВ со словесной разборчивостью человека в одних услови-

ях. Тестовые фразы озвучены людьми (носителями языка) разного пола, не имеющими дефектов речи. Все полученные аудиозаписи (речевой корпус), размечены и сохранены в формате *.wav без фильтрации (PCM) и обработки, имеют параметры: 16 бит, моно, 48кГц. Аудиозаписи шума или помехи, имеют одинаковые параметры звукозаписи с исходным речевым корпусом.

Способность распознавания системы голосового человеко-машинного взаимодействия оценивалось на основе словесной разборчивости речи ($W\%$), под которой понимаем относительное количество (в процентах) правильно понятых слов (в случае человека) и их же распознанных в контексте (в случае СГЧМВ).

Значение коэффициента сигнал/шум варьировалось от -30 дБ до 30 дБ. Эмпирически было выявлено, что при значении коэффициента сигнал/шум равном -30 дБ, словесная разборчивость ($W\%$) в обоих исследованиях близка к нулю как у человека, так и у СГЧМВ. Использовался общепринятый шаг значения коэффициента сигнал/шум 5 дБ [2, 4], результат приведен на рис.1.

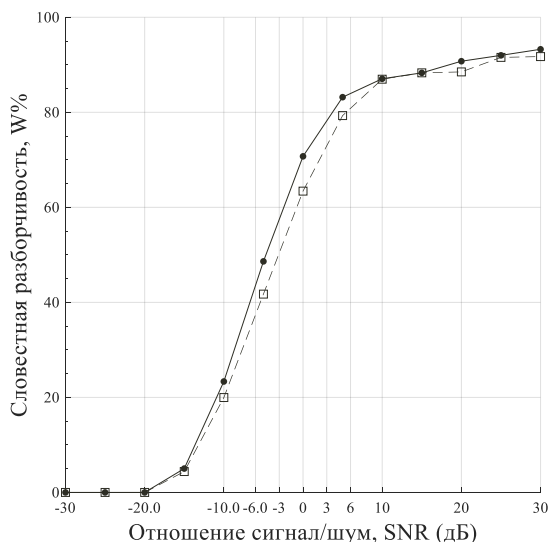


Рис.1. Оценка словесной разборчивости:
 —●— шумы в помещении (сплошная линия);
 -□- синтезированная помеха (пунктир)

Представленная зависимость с двумя аппроксимированными кривыми, получена в соответствии с методикой, описанной ГОСТ 16600-72.

Основываясь на критериях, описанных в работах [2, 4], опишем шкалу из пяти основных секторов, восприятия человеком, используемую для сопоставления с СГЧМВ:

- первый сектор – разборчивость $W \leq 10\%$, соответствует обнаружению факта речи;
- второй сектор $10\% < W \leq 20\%$, соответствует отсутствию возможности восприятия большей части слов;
- третий сектор $20\% < W \leq 30\%$, соответствует отсутствию возможности восприятия содержания текста;

- четвертый сектор $30\% < W \leq 40\%$, имеется возможность составления краткой семантической справки;

- пятый сектор $40\% < W \leq 100\%$ детальное содержание.

В свою очередь есть возможность разделить пятый (последний) сектор еще на три части относительно возможности человеком разобрать детальное содержание беседы [3, 4]. В промежутке от 40% и до 80% включительно человек может определить детальное содержание беседы только в результате многократного прослушивания, в промежутке от 80% до 90% включительно – разбор детального содержания осуществляется, не используя средств повторного прослушивания, а просто сконцентрировавшись и прислушавшись, а в промежутке от 90% до 100% любой человек может непреднамеренно детально разобрать суть.

Численные оценки исследования в соответствии с принятой шкалой, приведены в таблице 1.

Таблица 1. Словесная разборчивость СГЧМВ при использовании шума и помех на аудиозапись

| Словесная разборчивость, W% | Способ зашумления | |
|-----------------------------|-------------------------------------|---|
| | Интенсивность, шума в помещении, дБ | Интенсивность, синтезированной помехи, дБ |
| 10% | -13,40 | -12,80 |
| 20% | -10,20 | -09,60 |
| 30% | -08,00 | -07,40 |
| 40% | -05,50 | -05,20 |
| 80% | 04,30 | 04,80 |
| 90% | 20,90 | 21,50 |

В таблице 1 приведены значения разборчивости с делением на основную и дополнительную шкалу.

Первый сектор – $W\%$ составляет менее 10% при значении менее -20,00 до -13,40 дБ в случае наложения шума помещения и в случае использования синтезированной помехи от -30,00 до -12,80 дБ, что является на 0,60 дБ больше. В этом диапазоне человек ощущает только шум. Размерность первого сектора для первого способа зашумления составляет 16,60 дБ, для второго способа – 17,20 дБ.

Второй сектор – $W\%$ составляет менее 20%, но более 10% при значении от -13,40 до -10,20 дБ в случае наложения шума помещения и в случае использования синтезированной помехи от -12,80 до -9,60 дБ, что на 0,60 дБ больше. Размерность второго сектора для обоих способов зашумления составляет 3,20 дБ.

Третий сектор – $W\%$ составляет менее 30%, но более 20% при значении от -10,20 до -8,00 дБ в случае наложения шума помещения и в случае использования синтезированной помехи от -9,60 до -7,40 дБ, что на 0,60 дБ больше. Размерность третьего сектора для обоих способов зашумления соответствует 2,2 дБ.

Четвертый сектор – $W\%$ составляет менее 40%, но более 30% при значении от -8,00 до -5,50 дБ в

случае наложения шума помещения и в случае использования синтезированной помехи от -7,40 до -5,20 дБ, что на 0,30 дБ больше. Размерность четвертого сектора для первого типа (шум) зашумления составляет 2,50 дБ, для второго типа (помеха) – 2,20 дБ.

Пятый сектор – W% составляет менее 80%, но более 40% при значении от -5,50 до 4,30 дБ в случае наложения шума помещения и в случае использования синтезированной помехи от -5,20 до 4,80 дБ, что на 0,50 дБ больше. Размерность пятого сектора для первого способа зашумления составляет 9,80 дБ, для второго способа – 10,00 дБ.

Шестой сектор – W% составляет менее 90%, но более 80% при значении от 4,30 до 20,90 дБ в случае наложения шума помещения и в случае использования синтезированной помехи от 4,80 до 21,50 дБ, что на 0,60 дБ больше. Размерность шестого сектора для первого способа зашумления составляет 16,60 дБ, для второго способа – 16,80 дБ.

Седьмой сектор – W% составляет менее 80%, но более 40% при значении от 20,90 до 30,00 дБ в случае наложения шума помещения и в случае использования синтезированной помехи от 21,50 до 30,00 дБ. Размерность седьмого сектора в проведенном эксперименте для первого способа зашумления составила 9,10 дБ, для второго – 8,50 дБ.

Учитывая, что человек способен воспринять отличие в интенсивности сигнала при отношении сигнал/шум в 2 дБ [6], то начиная с четвертого сектора сопоставим способность распознавания СГЧМВ и словесной разборчивости речи человеком.

Транскрибированные аудиозаписи, содержащие все перечисленные выше воздействия, были прослушаны аудиторами (в возрасте до 20 лет). Прослушивание осуществлялось через Hi-res наушники (с типовыми параметрами ЦАП 48 кГц, 16 бит), результат усредненного восприятия синтезированной помехи (таблица 2):

Таблица 2. Восприятие речи аудитором, при использовании синтезированной помехи с интенсивностью от -30 дБ до 0 дБ

| № | Интенсивность | Восприятие акустической среды |
|---|---------------|--|
| 1 | -30 дБ | слышен шум |
| 2 | -25 дБ | слышно речь, но невозможно разобрать границы слов |
| 3 | -10 дБ | при многократном прослушивании, все слова определены |
| 4 | -5 дБ | при многократном прослушивании, правильно определены все слова |
| 5 | 0 дБ | все слова восприняты |

Результаты анализа человеком искаженных аудиозаписей добавление шума помещения. Основным свойством добавляемого шума, является монотон-

ность без четко выделяемых на слух голосов у записанных людей в помещении.

Таблица 3. Восприятие речи аудитором, при использовании естественного шума в помещении с интенсивностью от -30 дБ до 0 дБ

| № | Интенсивность | Восприятие акустической среды |
|---|---------------|--|
| 1 | -30 дБ | слышен шум |
| 2 | -25 дБ | слышно речь, но невозможно разобрать границы слов |
| 3 | -20 дБ | слышно речь, но невозможно разобрать слова |
| 4 | -15 дБ | при многократном прослушивании, все слова определены |
| 5 | 0 дБ | все восприняты без затруднения |

Анализируя полученные данные видно, что при использовании естественного шума человек начинает распознавать речь при отношении сигнал/шум -15 дБ, а при использовании синтезированного шума – при отношении сигнал/шум -10 дБ, что на 5 дБ меньше, чем при использовании естественного шума.

Заключение

Разница между распознаванием речи человека СГЧМВ незначительна, но хуже при воздействии синтезированной помехи (с равномерным законом распределения). Способность распознавания фраз системами ГЧМВ выше в условиях воздействия естественного шума.

Сопоставляя способность распознавания СГЧМВ и словесной разборчивости речи человеком отметим, что разборчивость речи человеком при добавлении синтезированной помехи лучше по сравнению с СГЧМВ и разница составляет 2 дБ. При добавлении естественного шума разница составляет 7,6 дБ.

Полученные результаты свидетельствуют о том, что возможно наличие шума, в условиях которого СГЧМВ будут не способны распознавать речь человека, в то время как человек сможет беспрепятственно воспринимать информацию из акустической среды.

Список литературы

- ГОСТ 16600-72. Передача речи по трактам радиосвязи [Текст]. – Введ. 1972-27-09. – М. : Изд-во стандартов, 1972. – 90 с.
- Дворянкин С. В., Макаров Ю. К., Хорев А. А. Обоснование критериев эффективности защиты речевой информации от утечки по техническим каналам //Защита информации. Инсайд. – 2007. – №. 2. – С. 18-25.
- ГОСТ Р 50840-95. Государственный стандарт Российской федерации. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. – М. : Госстандарт России, 1997. – 230 с.

4. Волчихина, М. В. Оценка уровня словесной разборчивости на границе контролируемой зоны офисного помещения артикуляционным методом / М. В. Волчихина // Вестник Воронежского института МВД России. – 2022. – № 2. – С. 105-112.

5. Информационная система оценки распределения внимания операторов в условиях группового управления / В. А. Аветисян, Е. А. Тарадайко, Н. А. Тузов // Информатика: проблемы, методы, технологии : Материалы XXIV Международной научно-практической конференции им. Э.К. Алгазина, Воронеж, 14–15 февраля 2024 года / Под редакцией Д.Н. Борисова. – С. 1162-1171.

6. Алдошина И. Основы психоакустики //Слух и речь. Часть. – 2002. – Т. 2. – С. 54-58.

АНАЛИЗ ЭФФЕКТИВНОСТИ АЛГОРИТМОВ СОГЛАСОВАНИЯ В ЗАДАЧАХ ФЕДЕРАТИВНОГО ДЕЦЕНТРАЛИЗОВАННОГО ОБУЧЕНИЯ ДЛЯ СЕТЕЙ С НЕНАДЁЖНЫМИ УЗЛАМИ

Гурковский А.А.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ*

Аннотация

В работе анализируется работа алгоритмов децентрализованной (многоагентной) оптимизации в условиях ненадежной работы агентов, т.е. в ситуации, когда агенты с положительной вероятностью могут быть недоступны для других агентов и не участвовать в оптимизации общей целевой функции, однако могут оптимизировать свою локальную целевую функцию. Исследуется влияние параметров модели и топологии сети на эффективность работы алгоритма.

Введение

Задачи децентрализованной оптимизации являются важной составной частью алгоритмов федеративного децентрализованного обучения. Федеративное обучение – это разновидность алгоритмов машинного обучения, в которых задача построения модели решается не на едином сервере, а отдельными пользователями (агентами). Преимущество таких алгоритмов в том, что при построении модели агенты обмениваются не своими локальными данными, а только вычисляемыми параметрами модели. Таким образом, предотвращается утечка данных, обеспечивается повышенная безопасность, а также экономится объём пересылаемой информации. Федеративное обучение включает в себя две основные разновидности алгоритмов – централизованные, характеризующиеся наличием центрального сервера, который агрегирует локальные параметры модели каждого из агентов, и децентрализованные (FDL), в которых отсутствует единый сервер, и агенты передают свои параметры модели другим агентам (в соответствии с

топологией сети) [1]. Под топологией сети понимается граф, который определяет порядок взаимодействия агентов (вершины графа соответствуют агентам, ребра соответствуют наличию коммуникаций между агентами) и коммуникационная матрица W , каждый элемент которой определяет вес взаимодействия между агентами. Заметим, что коммуникационная матрица полностью определяет структуру сети (и соответствующего графа), так как ее можно рассматривать как матрицу смежности взвешенного графа. Элемент $w_{ji} > 0$ тогда и только тогда, когда в графе есть ребро, соединяющее вершину j и вершину i . Обычно рассматривают стохастические матрицы W . Федеративные децентрализованные алгоритмы (FDL) имеют много преимуществ перед классическими алгоритмами машинного обучения [2], но, вместе с тем, при их применении возникают дополнительные проблемы. Одна из них – проблема согласования локальных параметров модели. Напомним, что исходная задача состоит в том, чтобы для выбранного типа математической модели найти единый набор оптимальных параметров на основе совокупности всех локальных данных, которыми обладают агенты. Но в соответствии с алгоритмами федеративного обучения задача оптимизации является децентрализованной, т.е. каждый агент, решает свою локальную задачу поиска оптимальных параметров модели, и найденные агентами оптимальные решения не обязаны совпадать. Поэтому возникает необходимость использовать алгоритмы согласования данных (алгоритмы консенсуса) [3]. Помимо проблемы согласования параметров модели в алгоритмах федеративного обучения необходимо учитывать переменную структуру связей агентов, которая может быть вызвана помехами при передаче информации, недостаточными вычислительными мощностями агентов, сбоями в работе и т.п. В [4] исследовалась скорость сходимости алгоритма в сети с медленно меняющимся по детерминированному правилу графом взаимодействия агентов. В [7] изучался алгоритм в ситуации, когда при передаче данные зашумляются. Предложен метод, который позволяет уменьшать рассогласование локальных параметров модели, т. е. улучшать сходимость алгоритма. В [8] анализируется сходимость децентрализованного алгоритма оптимизации, основанного на методе стохастического градиента, для некоторого класса сетей с переменной топологией. В [6] предложено обобщение метода Франка-Вульфа для решения задачи децентрализованной оптимизации для переменных сетей агентов. Рассмотрен случай как детерминированной последовательности графов, так и стохастической последовательности.

В данной работе мы рассматриваем задачу децентрализованной оптимизации в рамках децентрализованного алгоритма федеративного обучения в условиях ненадежной работы агентов. А именно, мы рассматриваем ситуацию, когда каждый агент, независимо от остальных, может стать недоступным, а именно, он не передает и не принимает никакую информацию от других агентов, однако продолжает решать свою локальную задачу оптимизации. Отме-

тим, что в [4] изучается работа алгоритма децентрализованной оптимизации в случае, когда последовательность коммуникационных матриц представляет собой случайную выборку из некоторого конечного набора дважды стохастических матриц с некоторыми дополнительными условиями на них. В данной работе мы рассматриваем более общую ситуацию, когда последовательность коммуникационных матриц образует эргодическую цепь Маркова. Мы показываем сходимость алгоритма децентрализованной оптимизации и анализируем зависимость и анализируем влияние параметров цепи Маркова на сходимость.

Задача децентрализованной оптимизации

Имеется система из n агентов. Агенты могут взаимодействовать и обмениваться информацией с соседями, которые определяются соответствующим графом или коммуникационной матрицей. Каждый агент знает свою целевую функцию $f_i(x)$, $x \in R^d$. Задача заключается в минимизации целевой функции, которая есть сумма целевых функций агентов

$$f(x) = \frac{1}{n} \sum_i f_i(x) \rightarrow \min$$

В общем виде алгоритм децентрализованной оптимизации работает следующим образом: в начальный момент каждый агент выбирает начальное состояние - начальное приближение решения своей локальной задачи $f_i(x) \rightarrow \min$;

1. Каждый агент делает шаг спуска согласно выбранному алгоритму оптимизации, получает промежуточную точку, и передает значение промежуточной точки своим соседям;

2. Каждый агент обновляет свое состояние в соответствии с выбранным алгоритмом согласования и получает следующую точку - приближение к решению задачи. После этого п.1 и 2 повторяются.

Применительно к алгоритмам федеративного децентрализованного обучения (FDL) описанная выше задача интерпретируется следующим образом. Каждый агент имеет доступный только ему набор данных (датасет). Локальная целевая функция $f_i(x)$ представляет собой локальную функцию потерь, вычисленную i -ым агентом по своему датасету. Задача состоит в том, чтобы по имеющимся у всех агентов данным построить наилучшую единую модель (общую для всех агентов), то есть, найти такой набор параметров x^* , который минимизирует суммарную функцию ошибок: $f(x^*) = \min_x [f(x) = \frac{1}{n} \sum_i f_i(x)]$. В результате решения задачи, состояния агентов, т. е. решения их локальных задач, должны совпадать. Для решения описанной выше задачи часто используют стохастический градиентный спуск (в качестве алгоритма оптимизации) и алгоритм согласования типа ДеГрута [6]. Таким образом, одна итерация данного децентрализованного федеративного алгоритма состоит из двух шагов. Сначала по элементам случайной подвыборки [5] каждый агент вычисляет стохастический градиент $g_i^{(t)}$ локальной функции f_i и осуществляет сдвиг в направ-

лении антиградиента: $x_i^{(t+1/2)} = x_i^{(t)} - n_t * g_i^{(t)}$. Эти вычисления могут производиться параллельно. За этим следует шаг согласования параметров, агент вычисляет свое новое состояние как средневзвешенное по соседям (соседи определяются коммуникационной матрицей):

$$x_i^{(t+1)} = \sum_j w_{ij}^{(t)} * x_j^{(t+1/2)}$$

Здесь $w_{ij}^{(t)}$ - элементы коммуникационной матрицы на итерации с номером t . В случае переменной коммуникационной матрицы нужно определить правило ее построения на каждой итерации). В ходе работы алгоритма отслеживается ошибка согласования $\frac{1}{n} \sum_i ||(x_i - x^*)||^2$, которая характеризует качество работы алгоритма. На рис.1. представлены графики ошибки согласования, которые получены в результате применения децентрализованного федеративного алгоритма для построения модели линейной регрессии на синтетических данных, для сети из 25 агентов, полным графом, и коммуникационной матрицей с элементами $w_{ij}^{(t)} = 1/n$. Параметр σ^2 здесь определяет оценку сверху дисперсии стохастического градиента. Мы наблюдаем существенное ухудшение работы алгоритма при увеличении дисперсии σ^2 стохастического градиента.

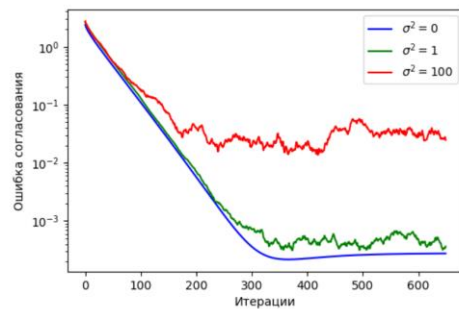


Рис.1. Ошибки согласования для модели линейной регрессии из 25 агентов

Модель с «ненадежными агентами»

В данной работе мы рассматриваем описанную выше оптимизационную задачу в условиях случайного отключения агентов. Каждый агент (независимо от других агентов) на каждой итерации может находиться в одном из двух положений: включён (on) и отключён (off). Отключённый агент не может принимать данные от других агентов, а также передавать свои параметры. Для отключённого агента j в матрице смежности все элементы w_{ij} и w_{ji} равны 0, за исключением элемента $w_{jj} = 1$ на диагонали, что соответствует обучению в автономном режиме. Для включенного элемента все w_{ij} равны $1/\text{степень узла } i$. Обозначим p_{off} - вероятность отключения на шаге $t+1$ при условии, что на шаге t агент включён, и p_{on} - вероятность включения на шаге $t+1$ при условии, что на шаге t агент выключен. Предположим, что p_{off} и p_{on} не зависят от номера шага и одинако-

вы для всех агентов. Таким образом, положение каждого агента описывается однородной эргодической цепью Маркова с двумя состояниями. Таким образом, в рассматриваемой модели последовательность коммуникационных матриц $\{W^{(t)}\}$ образует также однородную эргодическую цепь Маркова.

В работе построена имитационная модель сети с ненадежными агентами и реализован алгоритм децентрализованной оптимизации. Применим описанный в предыдущем разделе алгоритм для задачи определения параметров линейной регрессии для сети из 25 агентов на синтетических данных. На рис.2 приведены результаты моделирования для сети с надежными агентами и сети с ненадежными агентами.

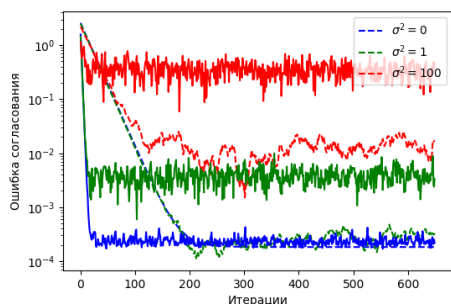


Рис.2. Графики ошибки согласования для сети с отключением агентов (непрерывная линия) и сети без отключения (пунктир) для полностью связной топологии. $p_{off} = 0.2$, $p_{on} = 0.8$

Результаты моделирования показывают, что для сети с ненадежными агентами ошибка согласования увеличивается по сравнению с ситуацией без отключений агентов. Отметим, что результаты, полученные для сети без отключения агентов, согласуются с результатами [6]-[7], что характеризует корректность работы реализованного алгоритма. На рис.3 приведены результаты моделирования при различных наборах вероятностей p_{off} и p_{on} . Как мы можем наблюдать, значение вероятности включения p_{on} агента влияет сильнее на итоговое значение ошибки согласования, чем значение вероятности его отключения p_{off} .

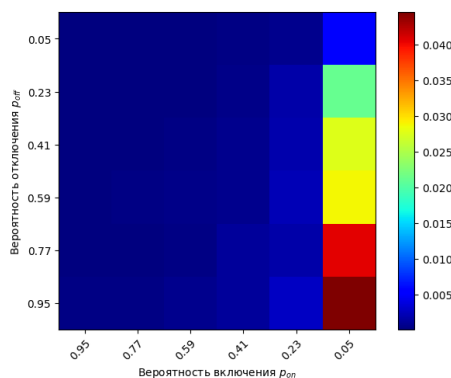


Рис.3. Ошибки согласования для полностью связной сети с отключением агентов для различных p_{off} и p_{on}

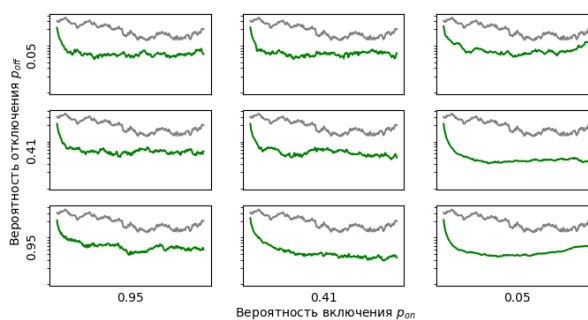


Рис.4. Ошибки согласования для сети с отключением агентов (зелёный) и без отключения агентов (серый)

Результаты моделирования выявили интересный эффект. Если при коммуникации присутствуют внешние помехи (реализованные в виде белого гауссовского шума), то можно наблюдать улучшение результатов согласования локальных параметров (рис.4) для сетей с малым числом связей. На рис. 4 приведены результаты моделирования для сети с графом кольцо для различных значений вероятностей p_{off} и p_{on} .

Заключение

В работе сформулирована задача децентрализованной (многоагентной) оптимизации общей функции потерь в условиях ненадёжной работы агентов. Для решения описанной выше задачи применили один из вариантов децентрализованного федеративного алгоритма, который сочетает стохастический градиентный спуск (в качестве алгоритма оптимизации) и алгоритм согласования типа ДеГрута. Проведен анализ работы алгоритма для двух топологий сети (полностью связный граф и граф-кольцо). Результаты моделирования показали, что алгоритм решает задачу децентрализованной оптимизации и в ситуации с ненадежными агентами. Выявлено, что эффективность алгоритма зависит от типа данных и вида графа. В дальнейшем исследовании планируется определить влияние параметров p_{off} и p_{on} на скорость сходимости алгоритма для различных графов сети.

Список литературы

1. Yuan, L, Sun, L, Yu, Ph, Wang, Z. (2023). Decentralized Federated Learning: A Survey and Perspective. – P. 1-6
2. X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, W. Zhang, and J. Liu, “Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent,” Advances in Neural Information Processing Systems, vol. 30, 2017 – P. 5-8
3. Proskurnikov, A, Tempo, R. (2017). A Tutorial on Modeling and Analysis of Dynamic Social Networks. Part I. Annual Reviews in Control. 43. – P. 2-8
4. Metelev, D, Beznosikov, A, Rogozin, A, Gasnikov, A, Proskurnikov, A. (2023). Decentralized optimization over slowly time-varying graphs: algorithms and lower bounds. Comput. Management Science. 21.– P. 5-21
5. G. Garrigos, R. M. Gower, “Handbook of Conver-

gence Theorems for (Stochastic) Gradient Methods,” January 27, 2023 <https://arxiv.org/abs/2301.11235> – P. 25-29

6. A. Koloskova, N. Loizou, S. Boreiri, M. Jaggi, and S. Stich, “A unified theory of decentralized SGD with changing topology and local updates,” in International Conference on Machine Learning, pp. 5381–5393, PMLR, 2020. – P. 6-9

7. P. Chellapandi, A. Upadhyay, A. Hashemi and S. H. Žak, "On the Convergence of Decentralized Federated Learning Under Imperfect Information Sharing," in *IEEE Control Systems Letters*, vol. 7, pp. 2982-2987, 2023 – P. 4-6

8. Ye, Hao & Liang, Le & Li, Geoffrey. (2022). Decentralized Federated Learning With Unreliable Communications. *IEEE Journal of Selected Topics in Signal Processing*. 16.– P. 6-9

РАЗРАБОТКА СИСТЕМЫ ПРЕДСКАЗАНИЯ ПАРАМЕТРОВ ЦИФРОВЫХ СХЕМ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

*Маркелов М.С.¹, Сорочайкин А.Я.³,
Подгорный Л.Е.², Лоскутов М.И.²*

*Национальный исследовательский университет
«Высшая школа экономики»,*

¹департамент электронной инженерии,

²департамент компьютерной инженерии,

³департамент прикладной математики

МИЭМ НИУ ВШЭ

Аннотации

Основная цель данной работы заключается в разработке программного модуля САПР для предсказания параметров комбинационных схем на маршруте проектирования ASIC с использованием методов машинного обучения. После загрузки в программу цифровой схемы в виде Verilog файла, будут рассчитаны такие характеристики схемы как площадь, энергопотребление, задержка сигнала.

Введение

Быстрая и точная оценка параметров комбинационных схем требуется для оптимизации процесса проектирования СБИС [1]. Весь процесс, от RTL до GDSII, требует различных оптимизаций и расчетов промежуточных характеристик схемы. При оценке схемы необходимо учитывать задержку сигнала, потребление энергии, площадь чипа и другие параметры. Однако, широко используемые методы расчета этих параметров на данный момент требуют больших вычислительных мощностей и занимают много времени. На данный момент, для расчета характеристик СБИС на этапе проектирования, уже используются методы машинного обучения [2], но используются этим, пока что, только иностранные компании, программный код которых находится в закрытом доступе.

Представление комбинационной схемы

Датасет с исходными схемами и их параметрами (площадь схемы в мкм, задержка сигнала в нс), с использованием которого будут обучены и протестированы алгоритмы, представлен в формате Verilog файлов. Стоит отметить, что представленные в датасете схемы были предварительно оптимизированы и наиболее приближены к финальной стадии перед технологическим отображением (technology mapping [3]).

Одной из важных задач данной работы является выбор и реализация представления Verilog файла (рис.1) в оптимальном для обучения виде. По результатам исследования было принято решение остановиться на двух форматах: представление схемы в виде графа и в виде эмбедингов [4].

```
module RandLevel(
    input x0,
    output f1, f2, f3, f4, f5
);

    wire wr54, wr55, wr56, wr57, wr58, wr59;

    nor (wr54, x0);
    xor (wr55, x0);
    or (wr56, wr55);
    xor (wr57, wr54);
    nor (wr58, wr57);
    nand (wr59, wr56, wr57);
    assign f1 = wr59;
    assign f2 = wr59;
    assign f3 = wr58;
    assign f4 = wr58;
    assign f5 = wr58;
endmodule
```

Рис.1. Цифровая схема в формате Verilog

Графовое представление необходимо для использования графовых нейронных сетей [5]. Для перехода от формата Verilog файла к графу используется программа, разработанная в рамках предыдущей итерации исследования. Графическая визуализация результата работы данной программы представлена на рис.2.

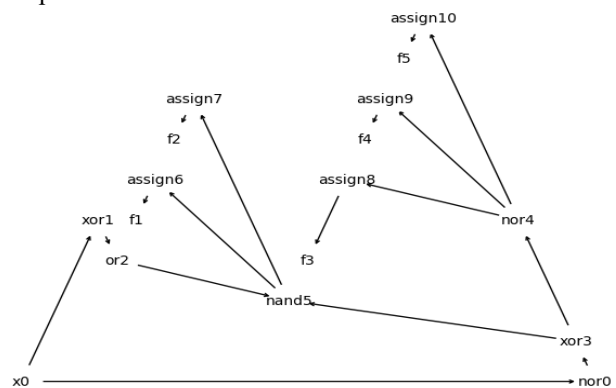


Рис.2. Визуализация графового представления
цифровой схемы

Чтобы иметь возможность обучать другие алгоритмы и нейронные сети, была разработана про-

грамма для представления схемы в виде эмбедингов [6]. В таком виде каждому элементу цифровой схемы (логическому вентилю, входу, выходу) будет соответствовать числовой вектор и в итоге схема будет описываться матрицей (рис.3), которая будет подана на вход разработанным алгоритмам. При разработке этой программы использовался алгоритм Node2Vec [7].

```

1 17 4
2 nand5 -0.014264016 0.0067821136 0.1274588 0.22600007
3 f1 -0.23287623 -0.17768417 0.16142984 0.22445577
4 xor3 -0.1262822 -0.09273814 0.18369053 -0.037776984
5 or2 -0.1138487 0.16437571 -0.12164114 -0.04479246
6 f3 0.07188494 0.02484981 -0.20712799 -0.23609473
7 f5 0.18266171 0.1268239 0.16885886 0.019204922
8 f2 0.15856756 -0.08504516 -0.02385313 0.14437266
9 assign7 -0.18813507 -0.09833961 -0.1880952 -0.022821397
10 assign6 0.23846479 -0.18301138 -0.058307774 -0.048345175
11 f4 0.20203333 -0.14800903 0.00075879414 -0.11860531
12 nor4 -0.240737 0.12592328 -0.21944775 -0.10892028
13 assign8 -0.0009714267 -0.0073517007 -0.19151205 0.2403359
14 nor0 0.12374811 0.23197035 -0.20451231 0.113604024
15 xor1 -0.10381037 0.021074863 0.21238047 -0.11113332
16 assign10 0.11270727 -0.16919261 -0.088808894 0.23545338
17 assign9 -0.03927673 0.008304275 -0.10396229 -0.1918727
18 x0 -0.038456403 0.06220409 -0.022067491 0.13862616
19

```

Рис.3. Представление схемы в виде эмбедингов

Первые этапы обучения стандартных алгоритмов

Следующий этап проекта заключается в обучении нескольких алгоритмов и нейронных сетей. На данный момент было обучено несколько стандартных моделей, на датасете небольшого объема [8]. В данной программе были предсказаны только площадь и задержка цифровых схем, далее планируется добавить энергопотребление.

Исходный датасет был представлен в формате эмбедингов размерности 25, объем датасета составил 13824 схемы.

При анализе гистограмм распределения предсказываемых параметров, представленных на рис.4, было выявлено и убрано небольшое количество выбросов и некорректных данных, после чего размер датасета сократился до 12844 схем.

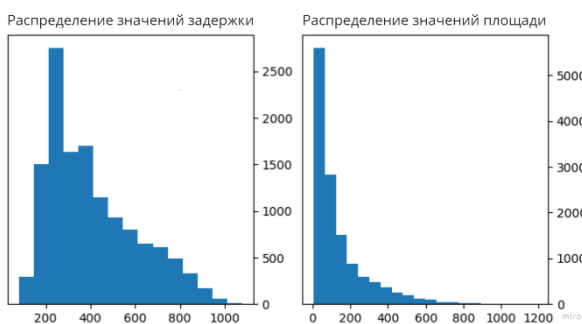


Рис.4. Гистограммы с распределением величин площади и задержки цифровых схем

При приведении данных к подходящему для обучения выбранных моделей виду (каждая схема должна быть представлена вектором одинаковой размерности) возникла следующая проблема: каждый узел

схемы представлен вектором длины 25, но количество узлов у схем разное, следовательно, размерности итоговых векторов будут различны для разных схем. В качестве решения выявленных проблем были реализованы следующие алгоритмы:

- дополнение всех векторов нулями до наибольшей, среди имеющихся в датасете векторов, размерности;
- дополнение всех векторов средними значениями по всей выборке до наибольшей, среди имеющихся в датасете векторов, размерности.

Различий в точности предсказаний параметров в зависимости от выбора решения этой проблемы выявлено не было.

На данный момент были обучены следующие алгоритмы: градиентный бустинг (GBM), полносвязная сверточная нейронная сеть (CNN). Для оценки качества обученных моделей использовались следующие метрики коэффициент детерминации [9] R^2 , средняя абсолютная ошибка (MAE), относительная ошибка (RE). Результаты обучения представлены в Таблице 1.

Таблица 1. Метрики качества моделей

| Алгоритм | Параметр схемы | Название метрики | Значение метрики |
|----------|----------------------|------------------|------------------|
| GBM | Площадь, мкм | R^2 | 0.983 |
| GBM | Площадь, мкм | MAE | 12.423 |
| GBM | Площадь, мкм | RE | 0.101 |
| CNN | Площадь, мкм | RE | 0.136 |
| GBM | Задержка сигнала, нс | R^2 | 0.982 |
| GBM | Задержка сигнала, нс | MAE | 12.392 |
| GBM | Задержка сигнала, нс | RE | 0.100 |
| CNN | Задержка сигнала, нс | RE | 0.123 |

Заключение

В ходе работы были выбраны и реализованы, наиболее подходящие для машинного обучения, варианты представления цифровых схем. Были обучены распространенные алгоритмы (градиентный бустинг, полносвязная сверточная нейронная сеть).

Анализируя значения, выбранных для оценки обученных алгоритмов, метрик, можно сделать вывод о довольно высокой точности предсказаний обученных моделей. В дальнейшем будет произведена дополнительная работа по обучению других ML алгоритмов и нейронных сетей таких, как: CatBoost, случайный лес (Random Forest Regressor), нейронная сеть с механизмом самовнимания (SelfAttentionNN).

Список литературы

1. S. L. Harris and D. M. Harris, Digital Design and Computer Architecture // RISC-V Edition. Morgan Kaufmann – 2021.
2. Vidya A. Chhabria et. al. MAVIREC: ML-Aided Vectors IR-Drop Estimation and Classification // arXiv:2012.10597v1 – 2020.
3. G. De Micheli, Technology Mapping of Digital Circuits // Proceedings, Advanced Computer Technology, Reliable Systems and Applications – 1991.
4. Эмбединги признаков и повышение точности ML-моделей. [Электронный ресурс]. URL: <https://habr.com/ru/companies/wunderfund/articles/590651/> (Дата обращения 10.02.2024).
5. F. Scarselli et. al. The Graph Neural Network Model // IEEE Trans. Neural Netw. – 2009.
6. Программа для перевода verilog файла в эмбединги [Электронный ресурс]. URL: https://git.miem.hse.ru/1799/verilog_to_embendings/-/blob/master/many_files.ipynb (Дата обращения 19.02.2024).
7. A. Grover, J. Leskovec. node2vec: Scalable Feature Learning for Networks // arXiv:1607.00653v1 – 2016
8. Программный код с обученными моделями [Электронный ресурс]. URL: https://git.miem.hse.ru/1799/verilog_to_embendings/-/blob/master/ml_models_with_optimathed_circuits.ipynb (Дата обращения 19.02.2024).
9. Коэффициент детерминации. [Электронный ресурс]. URL: <https://wiki.loginom.ru/articles/coefficient-of-determination.html> (Дата обращения 19.02.2024).

СТАТИСТИЧЕСКИЙ АРБИТРАЖ НА СВЕРХКРАТКОСРОЧНОМ ГОРИЗОНТЕ С ИСПОЛЬЗОВАНИЕМ ПОКАЗАТЕЛЯ ХЕРСТА

Кожмякин Д.А.

*Самарский государственный
экономический университет,
кафедра статистики и эконометрики
Института национальной и мировой экономики*

Аннотация

Показатель Херста — это важная характеристика временного ряда, позволяющая определить степень его персистентности. В работе рассматривается применение показателя Херста для высокочастотной торговли на бирже. Данный показатель может быть использован при отборе активов и подборе весов для построения портфеля статистического арбитража. В

итоге приводится общая схема алгоритма стратегии статистического арбитража на основе показателя Херста.

Введение

Методы анализа временных рядов являются неотъемлемой частью арсенала биржевых спекулянтов. В отличие от случайного блуждания, цены финансовых активов могут содержать скрытые закономерности. Особенно явно проявляется свойство долгосрочной памяти на временных рядах синтетически сконструированных инструментов. Такие инструменты обычно обладают антиперсистентной природой, что позволяет успешно применять на них контртрендовые стратегии. Данный подход называется статистическим арбитражем.

Целью текущего исследования является представление подхода к построению стратегии статистического арбитража, которая может быть реализована в рамках высокочастотной торговли. Соответственно, в исследовании стояли следующие задачи: изучение меры антиперсистентности временного ряда для нахождения активов и конструирования спреда, а также построение схемы алгоритма торговой стратегии на базе этого.

Для рассмотрения использовались данные криптовалютных активов ввиду их удобства и доступности.

В первом разделе исследования рассматривается показатель Херста, как мера долгосрочной памяти временного ряда.

Во втором разделе приведена общая схема построения стратегии статистического арбитража с использованием показателя Херста.

Исследование показателя Херста для биржевой торговли

Показатель Херста, H — хорошо известная мера, позволяющая оценить характеристики долгосрочной памяти временного ряда, такие как персистентность и антиперсистентность (склонность продолжать предыдущее движение и склонность менять направление). Также показатель связан с фрактальной размерностью временного ряда. Впервые данный показатель был предложен американским гидрологом Херстом, который работал над проблемой анализа многолетних колебаний годового стока рек, и выявил зависимость текущих колебаний от колебаний прошлых лет (эмпирическое значение показателя H составило 0,73) [1]. Значение H принято интерпретировать следующим образом:

- $H = 0,5$ — случайное блуждание: изменения происходят независимо;
- $1 > H > 0,5$ — персистентный временной ряд: текущее движение зависит от прошлого и имеет тенденцию продолжиться в ту же сторону;
- $0 < H < 0,5$ — антиперсистентный временной ряд: текущее движение зависит от прошлого и более склонно изменить направление.

Существуют различные методы расчета показателя Херста. В данной работе используется

метод, описанный в книге Майкла Л. Холлса Мура [2], он основан на оценке дисперсий разностей n -го порядка исходного временного ряда в увеличивающемся окне. После того как оценки дисперсий для разных порядков были получены, на логарифмированных данных (оценки дисперсий и порядки n) строится линейная модель, через коэффициент наклона линейной регрессии в таком случае и будет выражен искомым показатель. Данный метод был модифицирован путем применения формулы выборочной дисперсии, когда делитель корректируется на величину порядка n . Это позволило улучшить точность оценки параметра Херста, что видно на симуляции для случайного блуждания: с увеличением порядка n , оценка параметра становится заниженной, что вероятно, связано с уменьшением количества наблюдений при оценке дисперсии на разностях более старших порядков. Подобная корректировка расчета дисперсии помогает несколько снизить данный эффект, что позволяет использовать меньше наблюдений (см. рис.1).

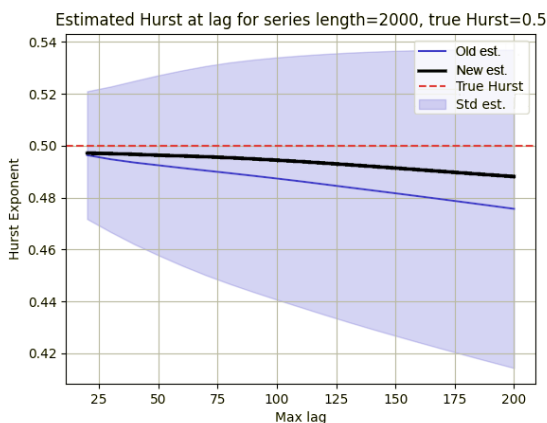


Рис.1. Симуляция оценки параметра Херста для случайного блуждания

Благодаря своей простоте и универсальности параметр Херста стал широко применяться за пределами гидрологии, так, анализ на основе параметра Херста используется в дендрохронологии, а также при изучении микросейсмической активности [3].

В практике биржевой торговли параметр Херста может быть использован как индикатор наличия возможностей для извлечения прибыли или неэффективности рынка [4]. Чаще всего параметр Херста используется, чтобы определить, в каком состоянии находится рынок: трендовом, контртрендовом или случайном. Гораздо интереснее с точки зрения свойств временного ряда рассматривать не цены одного актива, а некий синтетический инструмент, как комбинацию из нескольких активов. В сущности, такой инструмент представляет собой портфель, в котором по активам занимают как короткие, так и длинные позиции с разными весами (на практике короткие позиции не

всегда возможны и не всегда обязательны). Как правило, подобные синтетические инструменты обладают антиперсистентной природой, и хорошо подходят для контртрендовых торговых стратегий. Данный подход и соответствующие стратегии называют статистическим арбитражем или, если говорить более точно, — парным трейдингом или же баскет-трейдингом (торговля корзиной инструментов), а сам временной ряд синтетического инструмента — спредом.

На рис.2 представлен спред криптовалютных активов: $-1*ADAUSDT + 10*AVAXUSDT - 7*BTCUSDT - 5*DOTUSDT$. Стоит заметить, что при оценке показателя Херста использовались лаги в диапазоне 2-30 секунд. В работе использовались высокочастотные данные крипто-активов ввиду их удобства и доступности: рынок работает непрерывно, а секундные данные с биржи Binance — бесплатны.

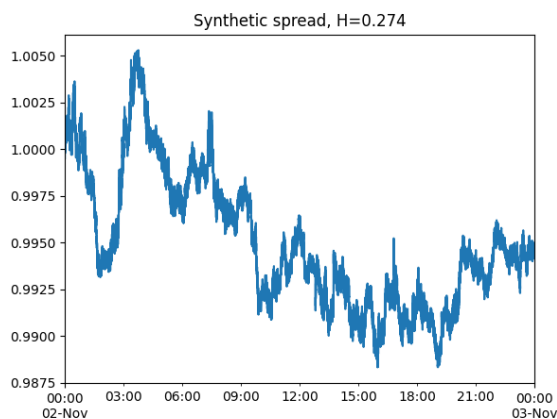


Рис.2. Спред $-1*ADAUSDT + 10*AVAXUSDT - 7*BTCUSDT - 5*DOTUSDT$

В парном трейдинге показатель Херста часто используется для отбора подходящих инструментов, как альтернатива подходам на основе минимального расстояния, корреляции или коинтеграции [5] [6]. Комбинация активов и их весов в портфеле считается наилучшей в таком случае, если временной ряд портфеля имеет наименьшее значение параметра Херста. Подобный портфель будет обладать хорошей прогнозируемостью, что позволит успешно торговать им как синтетическим активом.

Примечательно, что вычисление параметра Херста также может быть полезно в целях анализа привлекательных торговых возможностей и выбора ниши. На рис.3 видно, что на интервале до 30 секунд (примерно 2 в степени 5) существуют сильные контртрендовые возможности, и участники с достаточно скоростной инфраструктурой могли бы их эксплуатировать. После 30 секунд преимущество значительно снижается, о чем свидетельствует изменившийся наклон линии фактических данных.

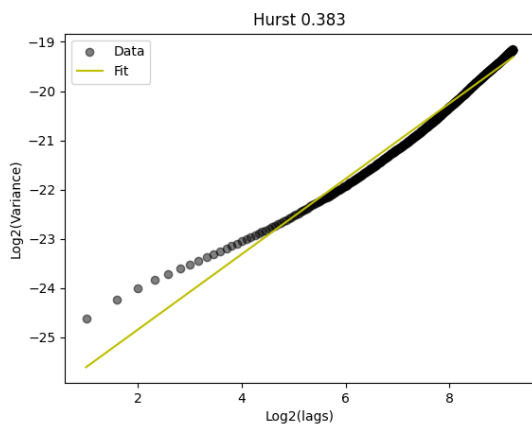


Рис.3. Структура дисперсии разностей спреда

Построение высокочастотной торговой стратегии арбитража

Для небольших спекулянтов, но обладающих достаточной инфраструктурой, наибольший интерес представляют стратегии высокочастотной торговли. Существует несколько основных видов и подкатегорий данных стратегий, среди которых статистический арбитраж [7].

Статистический арбитраж синтетических инструментов является одной из хорошо изученных количественных стратегий. Краусс выделяет несколько больших групп подходов, встречающихся в литературе, таких как методы измерения расстояния, подход на основе коинтеграции, методы временных рядов, стохастические методы и «другое» [8].

Не менее важно общее описание алгоритма статистического арбитража. Хи и др. приводят свою схему проведения исследования, которая во многом соотносится с общими шагами по тестированию торговой стратегии (см. рис.4) [9].

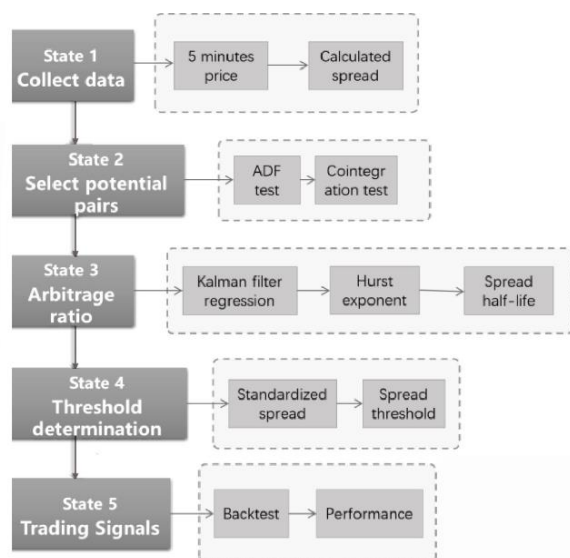


Рис.4. Схема исследования [9]

Построение стратегии статистического арбитража с использованием показателя Херста выглядит следующим образом:

1. Выбрать подмножество активов, удовлетворяющих определенным критериям торговли (например, уровню ликвидности или величине bid/ask спреда).

а) Чтобы сократить возможные варианты, можно дополнительно среди них выбрать n лучших активов (например, с высокой корреляцией или низкой дистанцией).

2. Сформировать все возможные корзины из m активов, перебрав каждый актив с целыми весами w_1, \dots, w_m , включая 0. Измерить показатель Херста.

а) Вместо метода полного перебора можно использовать альтернативный алгоритм.

3. Отобрать среди полученных спредов такие, величина оценки показателя Херста которых не превышает заданный порог.

4. Оптимизировать параметры контртрендового индикатора технического анализа на отобранных спредах, и если характеристики кривой доходности после издержек удовлетворяют заданным критериям, то запуск алгоритма с полученными параметрами на следующем периоде.

а) Вместо индикатора технического анализа можно использовать прогноз модели временных рядов.

Здесь длина окна для расчета показателя Херста, а также длина окна для оптимизации индикатора — являются отдельными параметрами для оптимизации. Также подмножество активов может периодически пересматриваться.

Заключение

Показатель Херста может быть взят за основу для построения стратегии статистического арбитража. Также с его помощью можно проанализировать наличие торговых возможностей на тех или иных временных интервалах. Для построения высокочастотной стратегии крайне важно использовать простые и эффективные алгоритмы в целях снижения задержек. Показатель Херста вполне удовлетворяет этим требованиям, но нужна альтернатива алгоритму полного перебора. В дальнейшем планируется использовать приведенную схему для оптимизации и тестирования торговой стратегии.

Список литературы

1. Христофоров А. В. Теория случайных процессов в гидрологии // М.: Изд-во МГУ. – 1994. – С. 47.
2. Michael L. Halls Moore. Successful Algorithmic Trading. Applying the scientific method for profitable trading results. (2015). – С. 90.
3. Любушин А. А. Анализ данных систем геофизического и экологического мониторинга. – Наука, 2007. – С. 60-63, С. 169.
4. Kyriazis N. A. A survey on efficiency and profitable trading opportunities in cryptocurrency

markets //Journal of Risk and Financial Management. – 2019. – Т. 12. – №. 2. – С. 2-5.

5. Bui Q., Ślepaczuk R. Applying Hurst Exponent in pair trading strategies on Nasdaq 100 index //Physica A: Statistical Mechanics and its Applications. – 2022. – Т. 592. – С. 126784.

6. Ramos-Requena J. P., Trinidad-Segovia J. E., Sánchez-Granero M. A. Introducing Hurst exponent in pair trading //Physica A: statistical mechanics and its applications. – 2017. – Т. 488. – С. 39-45.

7. Перстенева, Н. П. Классификация стратегий высокочастотной торговли / Н. П. Перстенева, Д. А. Кожемякин. – Текст : электронный // Экономика и предпринимательство. -2022. -№8(145). -С. 832-837.

8. Krauss C. Statistical arbitrage pairs trading strategies: Review and outlook //Journal of Economic Surveys. – 2017. – Т. 31. – №. 2. – С. 513-545.

9. He C. et al. An innovative high-frequency statistical arbitrage in Chinese futures market //Journal of Innovation & Knowledge. – 2023. – Т. 8. – №. 4. – С. 100429.

ОБЗОР МОДЕЛЕЙ НЕЙРОННЫХ СЕТЕЙ ДЛЯ УЛУЧШЕНИЯ КАЧЕСТВА РЕЧИ

Горбунцов А.А., Миркин Д.Е.

Национальный исследовательский университет

«Высшая школа экономики»,

департамент электронной инженерии

МИЭМ НИУ ВШЭ

Аннотация

Улучшение качества и общая очистка аудиозаписи голоса от шумов с помощью нейронных сетей является важной задачей области работы со звуком. Данный обзор раскрывает два основных подхода к решению данной задачи — улучшение качества аудиофайла и улучшение качества мел-спектрограммы. Рассмотрены основные архитектуры, методы и метрики эффективности этих моделей, с акцентом на повышении качества восприятия речевой информации. Отсутствие обсуждения вычислительных ресурсов и сложности обучения позволяет сконцентрироваться на анализе результатов, призванных максимально улучшить аудиофайлы для оптимального восприятия речи. Сравнивается эффективность моделей, а также выделяется модель, наиболее успешно справляющаяся с задачей.

Введение

Машинное обучение сегодня является неотъемлемой частью задач обработки звука. Современные технологии в области обработки речи стремительно развиваются, открывая новые перспективы для улучшения качества аудио-контента. Одним из ключевых направлений в этой области является применение нейронных сетей для очистки от шума и общего повышения качества восприятия речи. Тихую или невнятную речь, смешанную со звуками окружающей среды сложно слушать, из нее тяжело извлекать

смысл. В данной статье мы обзорно рассмотрим два основных типа моделей, предназначенных для улучшения качества речи: модели, направленные на обработку аудиофайлов, и те, которые работают с мел-спектрограммами.

Будет уделено внимание архитектуре моделей и основным метрикам их эффективности. В данной статье внимание сосредоточено исключительно на качестве восприятия речи, а вопросы сложности обучения и вычислительных ресурсов рассматриваться не будут, поскольку наша цель – проанализировать степень улучшения аудиофайлов для максимального повышения восприятия речевой информации. Также будет отмечена наиболее эффективная модель каждого типа.

Модели улучшения качества аудиофайлов

Для решения задачи улучшения качества сгенерированных записей голоса (speech enhancement) был проведен анализ существующих эффективности различных существующих на сегодня архитектур нейронных сетей на основе конкретных моделей. Рассмотренные модели обучались и валидировались на датасете VoiceBank-DEMAND - датасет, состоящий из английской речи, искаженной шумом окружающей среды [1]. Речевые файлы имеют длину от 3 до 10 секунд. Набор данных состоит из пар чистой и шумной речи от 28 разных носителей (14 мужчин, 14 женщин) говорящих по-английски. Чистая речь смешивается при соотношении сигнал/шум 0, 5, 10 и 15 дБ с различными шумами окружающей среды. Тестовый набор состоит из 824 от двух дополнительных спикеров (один мужской, один женский), которые не появляются в обучающей выборке, смешанных с отношением сигнал/шум 2,5, 7,5, 12,5 и 17,5 дБ с различными шумами окружающей среды. Все аудио имеет частоту дискретизации 16000 Гц и имеет формат WAV. Обучающая выборка содержит 11572 пары чистой и шумной речи, валидационная выборка состоит из 824 пар чистой и шумной речи.

Были рассмотрены шесть моделей различных архитектур. Рассмотрим каждую подробнее.

SEGAN (Speech Enhancement Generative Adversarial Network) использует адаптированную генеративно-состязательную сеть (GAN), представляющую собой соревнующиеся генератор и дискриминатор [2]. Генератор стремится улучшить качество речевого сигнала, а дискриминатор обучается различать оригинальный сигнал от улучшенного, модель может обучаться без парных данных.

CDiffuSE (Cascaded Diffusion-Enhanced Speech Enhancement) использует каскад диффузионных слов, которые позволяют моделировать распространение информации через графы или пространственные сети, для последовательного увеличения разрешения аудиосигнала [3]. Каждый слой обрабатывает сигнал с более высоким разрешением, модель эффективно управляет различными уровнями шума и обеспечивает последовательное улучшение качества сигнала.

MetricGAN ориентирована на улучшение оценок качества аудиосигнала (метрика) качества речи с учетом восприятия человека с использованием генера-

тивно-состязательной архитектуры [4]. Генератор и дискриминатор обучены оптимизировать конкретные метрики и точно оценивать градиенты, направляющие обучение. Модель фокусируется на улучшении спектральных и временных характеристик сигнала для повышения оценок качества.

CNN-BLSTM (Convolutional Neural Network с Bidirectional Long Short-Term Memory) объединяет сверточные нейронные сети (CNN) для извлечения пространственных признаков и двунаправленные рекуррентные нейронные сети с долгосрочной и краткосрочной памятью (BLSTM) для учета временных зависимостей в данных [5]. Модель способна учитывать и пространственные, и временные зависимости в аудиосигналах.

T-GSA (Temporal Generative Self-Attention) использует архитектуру трансформера, основанный на механизме гауссово-взвешенного само-внимания, веса внимания ослабляются в соответствии с расстоянием между целевыми и контекстными символами для обработки временных зависимостей в аудиосигналах. Позволяет более эффективно улавливать долгосрочные зависимости и динамику аудиосигналов с использованием механизма самовнимания [6].

Deep Complex U-Net расширяет архитектуру U-Net, которая состоит из кодировщика и декодировщика, соединенных "U"-образной связью [7]. Кодировщик уменьшает размерность данных, а декодировщик восстанавливает исходный размер. Имеет связи, обеспечивающие передачу информации между кодировщиком и декодировщиком, с использованием комплексных чисел для представления аудиосигналов. Все сверточные слои UNet заменяются сложными сверточными слоями, инициализированными в соответствии со специальными критериями для обработки аудиосигналов, чтобы предотвратить потерю пространственной информации, а ядра свертки устанавливаются независимыми друг от друга, инициализируя веса в виде унитарных матриц для лучшего обобщения и быстрого обучения. На этапе декодирования для восстановления размера входных данных используются сложные деконволюционные операции. Модель может учитывать и обрабатывать фазовую информацию, что важно для точной обработки аудиосигналов.

Для оценки качества моделей использовались следующие метрики, отражающие восприятие речи человеком: PESQ (Perceptual Evaluation of Speech Quality) – оценивает качество речи с учетом восприятия человека, сравнивая эталонный сигнал с тестовым [8]. Результатом сравнения входного и выходного сигналов является оценка качества связи, которая аналогична усредненной субъективной оценке MOS (Mean Opinion Score), определяемой группой слушателей-экспертов согласно спецификации МСЭ-Т P.800. Оценки PESQ калибруются с использованием базы данных оценок MOS. Аналогичным образом работают связанные составные (composite) метрики: CSIG (Signal) - оценивает интенсивность искажения сигнала; CBAK (Background) - оценивает интенсивность фонового шума; COVL (Overall) - предоставляет общую оценку качества записи речи [9]. Метри-

ки оцениваются количеством баллов от 1, что речь практически не воспринимается, до 5, что означает свободное восприятие речи. В таблице 1 приведено сравнение этих для метрик исходного искаженного сигнала и очищенного от шума сигнала с помощью различных моделей.

Таблица 1. Сравнение улучшения метрик после обработки моделями по сравнению с исходными данными

| Модель для обработки данных | PESQ | CSIG | CBAK | COVL |
|-----------------------------|------|------|------|------|
| Исходные зашумленные данные | 1.99 | 3.37 | 2.49 | 2.66 |
| SEGAN | 2.16 | 3.48 | 2.94 | 2.80 |
| CDiffuSE | 2.52 | 3.72 | 2.91 | 3.10 |
| MetricGAN | 2.86 | 3.99 | 3.18 | 3.42 |
| CNN-BLSTM | 3.01 | 4.09 | 3.54 | 3.55 |
| T-GSA | 3.06 | 4.18 | 3.59 | 3.62 |
| Deep Complex U-Net | 3.24 | 4.34 | 4.10 | 3.81 |

В таблице 1 все модели отсортированы по улучшению метрик. Видно, что SEGAN демонстрирует худшие показатели, не сильно отличающиеся от исходных шумных данных. Каждая следующая модель улучшает качество все больше и больше. Модели, учитывающие пространственные характеристики аудиозаписей значительно превосходят генеративные модели по всем метрикам. Deep Complex U-Net превосходит все остальные модели примерно по всем метрикам более чем на 20 пунктов, что свидетельствует о ее высокой точности и пригодности к применению в сложных задачах. Такая точность была получена благодаря тому, что удачную архитектуру U-Net целенаправленно доработали для обработки временных и пространственных характеристик.

Проведенный анализ позволил получить общее представление об эффективности различных архитектур нейронных сетей в задаче улучшения качества речи на примере конкретных моделей, а также выделить важность пространственно-временных характеристик аудиозаписей. Модель Deep Complex U-Net показала наилучшую производительность.

Модели инверсии мел-спектрограммы

Инверсия мел-спектрограммы – это процесс преобразования мел-спектрограммы обратно в аудиосигнал. Мел-спектрограмма представляет собой визуальное представление спектра частот аудиосигнала, преобразованное в мел-шкалу. Инверсия мел-спектрограммы играет важную роль в задачах улучшения качества аудио, таких как синтез речи, улучшение качества звука и создание музыки, позволяя восстанавливать аудиосигнал с высокой точностью и натуральностью.

Был проведен анализ пяти современных моделей, специализирующихся на инверсии мел-

спектрограммы: Hi-Fi GAN[10], MelGAN[11], WaveNet[12], WaveGlow[13] и BigVSAN[14]. Каждая из этих моделей предлагает уникальные подходы и технологии для решения задачи синтеза аудио, обладая своими сильными и слабыми сторонами.

Для сравнения качества моделей был использован набор данных VCTK, который состоит примерно из 44200 коротких аудиоклипов, произнесенных 109 носителями английского языка с различными акцентами. Общая продолжительность всех аудиоклипов составляет около 44 часов. Аудиоформат – 16-битный PCM с частотой дискретизации 44 кГц. В данных экспериментах частота дискретизации была снижена до 22 кГц. Для использования в качестве валидационной выборки случайным образом были выбраны девять дикторов, и все их аудиоклипы были удалены из обучающего набора данных. Таким образом 100 дикторов составляют обучающую, и 9 дикторов составляют тестовую выборку. Все аудиоклипы нормализовались, чтобы предотвратить влияние различий в громкости на оценщиков. Все модели обучались до 2,5 миллионов шагов.

HiFi-GAN: Эта модель использует генеративно-состязательные сети для создания реалистичных аудиосигналов. Она отличается наличием нескольких дискриминаторов, работающих на разных масштабах и с различными временными периодами. Генератор HiFi-GAN преобразует мел-спектрограммы в аудиосигналы, используя сверточные сети с множественными входами для улучшения качества и естественности звучания. Дискриминаторы обучаются распознавать различия между настоящими и синтезированными сигналами, что позволяет модели улучшать качество генерации звука.

BigVSAN: Модификация генеративно-состязательной сети, которая интегрирует SAN для оптимизации процесса обучения дискриминаторов. BigVSAN направлена на достижение высокой производительности в синтезе речи, включая более точное определение оптимальной проекции для различения реальных и сгенерированных данных. Это достигается за счет включения дополнительного слоя, который оптимизирует проекцию признаков в пространстве характеристик.

WaveNet: Авторегрессионная модель, использующая глубокие сверточные нейронные сети для по-символьной генерации аудио. Она производит высококачественное звучание за счет большой глубины и сложности модели, что обеспечивает детальное воспроизведение тонких нюансов звука. Однако из-за своей вычислительной сложности WaveNet часто требует больше времени для генерации звука.

WaveGlow: Модель, сочетающая принципы авторегрессионного и генеративно-состязательного обучения. WaveGlow использует flow-based подход, что позволяет параллельно генерировать аудиосигналы и ускорять процесс синтеза, при этом сохраняя высокое качество звука. Эта модель обучается с помощью максимизации правдоподобия, обеспечивая быструю и эффективную генерацию аудио.

MelGAN: Легковесная модель, ориентированная на скорость работы. MelGAN использует стандарт-

ную архитектуру GAN с упрощенной сетью дискриминаторов для быстрой инверсии мел-спектрограмм в аудиосигналы. Несмотря на свою легковесность, модель способна генерировать качественные аудио в режиме реального времени, что делает ее подходящей для применения во многих задачах, связанных с синтезом речи.

Для сравнения и оценки моделей использовалась метрика MOS. Метрика Mean Opinion Score (MOS) — это стандартная мера, используемая для оценки качества аудио и видео, в частности в области телекоммуникаций и аудио синтеза. В контексте синтеза речи MOS используется для субъективной оценки естественности и качества звучания синтезированной речи. Диапазон значений MOS варьируется от 1 до 5. Оценки MOS собираются путем прослушивания группой оценщиков аудио образцов и выставления ими оценок по указанной шкале. Это субъективная оценка, так как она основана на восприятии человека.

Сравнительные показатели по метрикам для каждой модели в сравнении с исходными чистыми данными приведены в следующей таблице.

Таблица 2. Сравнение качества улучшения данных с помощью моделей в сравнении с исходными чистыми данные

| Модель для обработки данных | MOS |
|-----------------------------|-------------|
| WaveNet (MoL) | 3.86 ± 0.14 |
| WaveGlow | 3.96 ± 0.13 |
| MelGAN | 3.09 ± 0.11 |
| HiFi-GAN V1 | 4.18 ± 0.08 |
| BigVSAN | 3.43 ± 2.04 |
| Исходные чистые данные | 4.40 ± 0.06 |

На основе проведенного анализа моделей Hi-Fi GAN и BigVSAN показали наилучшие результаты по многим метрикам, демонстрируя высокое качество звучания и естественность синтезированной речи. WaveNet выделяется выдающимся качеством звучания за счет высокой вычислительной сложности. WaveGlow и MelGAN предлагают компромисс между качеством и скоростью генерации, подходя для задач реального времени.

Заключение

Достижения в области машинного обучения открывают широкие перспективы в области обработки аудио. Развитие архитектур нейронных сетей демонстрирует значительное улучшение показателей улучшения качества речи. Рассмотренные два основных типа моделей – те, что направлены на обработку аудиофайлов, и те, что работают с мел-спектрограммами – представляют перспективные инструменты для очистки от шума и общего повышения четкости и выразительности речи. Подчеркнуты технологические аспекты, которые влияют на решение данной задачи. Существующие сегодня решения хо-

рошо справляются с задачей и предоставляют большие возможности для практического применения. Отмечено, что Deep Complex U-Net для аудиофайлов и генеративный HiFi-GAN для мел-спектрограмм являются наиболее эффективными на данный момент решениями. Выделенные модели, показавшие высокую эффективность, можно применять в различных задачах, в которых важно восприятие речи.

Список литературы

1. Valentini-Botinhao C. et al. Investigating RNN-based speech enhancement methods for noise-robust Text-to-Speech // ISCA Speech Synthesis Workshop – ISCA, 2016.
2. Pascual S. et al. SEGAN: Speech Enhancement Generative Adversarial Network // arXiv preprint arXiv:1703.09452v3. – 2017.
3. Lu Y. et al. Conditional Diffusion Probabilistic Model For Speech Enhancement // arXiv preprint arXiv:2202.05256v1. – 2022.
4. Fu S. et al. MetricGAN: Generative Adversarial Networks based Black-box Metric Scores Optimization for Speech Enhancement // arXiv preprint arXiv:1905.04874v1. – 2019.
5. Kim J. et al. End-to-End Multi-Task Denoising for Joint SDR and PESQ Optimization // arXiv preprint arXiv:1901.09146v4. – 2023.
6. Kim J. et al. T-GSA: Transformer with Gaussian-weighted self-attention for speech enhancement // arXiv preprint arXiv:1910.06762v3. – 2020.
7. Choi H. et al. Phase-aware speech enhancement with Deep Complex U-Net // arXiv preprint arXiv:1903.03107v2. – 2020.
8. Hu Y. et al. Evaluation of Objective Quality Measures for Speech Enhancement // IEEE Transactions on Audio Speech and Language Processing 16. – IEEE, 2008. – С. 229-238
9. Reddy C.. et al. DNSMOS P.835: A non-intrusive perceptual objective speech quality metric to evaluate noise suppressors // arXiv preprint arXiv:2110.01763v4. – 2022.
10. Kong, J., Kim, J., & Bae, J. (2020). Hifi-gan: Generative adversarial networks for efficient and high fidelity speech synthesis. *Advances in Neural Information Processing Systems*, 33, 17022–17033.
11. Kumar, K., Kumar, R., de Boissiere, T., Gestin, L., Teoh, W. Z., Sotelo, J., de Brebisson, A., Bengio, Y., & Courville, A. C. (2019). Melgan: Generative adversarial networks for conditional waveform synthesis. *Advances in Neural Information Processing Systems*, 32.
12. Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *ArXiv Preprint ArXiv:1609.03499*.
13. Prenger, R., Valle, R., & Catanzaro, B. (2019). Waveglow: A flow-based generative network for speech synthesis. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3617–3621.
14. Shibuya, T., Takida, Y., & Mitsufuji, Y. (2023). BigVSAN: Enhancing GAN-based Neural Vocoders with Slicing Adversarial Network. *ArXiv Preprint ArXiv:2309.02836*.

АВТОМАТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ ТИПА ПЛАСТИКА ДЛЯ 3-D ПЕЧАТИ ПРИ ПОМОЩИ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

Горбунцов А.А., Васильев Г.В.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотации

Процесс 3D-печати требует определения типа используемого пластика, который имеет свои уникальные физические свойства и параметры печати. Для обеспечения высокой точности и повышенного контроля качества при производстве высокоточных компонентов, необходимо правильно определить тип пластика, используемого при печати.

В данном проекте предлагается исследовать возможности применения нейросетевых технологий для автоматического определения типа пластика при 3D-печати. Для этого планируется разработать набор алгоритмов, которые будут способны определить тип пластика на основе его физических свойств и параметров печати.

Введение

В последние годы рынок 3D печати значительно вырос и стал доступным для более широкой аудитории. Однако, при 3D-печати из пластика могут возникать различные проблемы, которые могут повлиять на качество печати и окончательный результат. Некоторые из распространенных проблем включают: отбегание, скручивание, растрескивание. Неадекватная адгезия: Если не будет достаточной адгезии между пластиком и печатным столом, объект может сдвигаться или отрываться во время печати. Поддержка: Некоторые модели требуют элементы поддержки, чтобы обеспечить правильное формирование сложных структур или свесов. Слоистость: 3D-печать происходит через постепенное наложение слоев пластика. Иногда это может приводить к заметной слоистости печати, особенно при недостаточной разрешающей способности или при использовании неправильной настройки. Температурные проблемы: печать из пластика требует соответствующей определенному типу материала температуры расплавления, чтобы обеспечить оптимальные результаты. Если температура слишком высокая или низкая, это может привести к проблемам с адгезией, отбеганию или отклонению от заданной формы объекта. Чтобы избежать этих проблем, важно правильно настроить 3D-принтер, используя рекомендации и настройки, предоставленные производителем пластика и 3D-принтера.

Для решения описанных выше проблем поставлена задача разработать систему, способную определять тип и характеристики пластика перед 3D печатью на основе программно-аппаратного комплекса. Для этого необходимо создать нейросетевые алгоритмы, обученные на большом наборе данных, содержащем информацию о физических свойствах и

параметрах печати для различных типов пластика. Целью является автоматическое определение типа пластика в процессе 3D-печати.

Программная часть решения

Чтобы определить тип пластика, необходимо идентифицировать его по набору уникальных свойств – параметров, а также определение типа возможно по увеличенной фотографии. В связи с этим поставлены две задачи – создание алгоритма определения типа по параметрам, и определение типа по фотографии. Были взяты параметры, которые можно измерить с помощью предлагаемого далее аппаратного комплекса. Для определения типа материала по параметрам предложен алгоритм, реализованный на языке Python. На основании информации, собранной на сайтах основных компаний-поставщиков пластиков для 3-Д печати была собрана таблица параметров для наиболее часто применяемых типах пластика. Таблица представлена на рис.1.

| Характеристики | ABS | SBS | PLA | Nylon | PETG | Flex |
|-------------------------------------|-----------|------------|-----------------|------------|-----------|-----------|
| Температура плавления | 175-210°C | 190-210°C | 175-180°C | 215-220°C | 222-225°C | 200-210°C |
| Температура размягчения | 100°C | 76°C | 50°C | 120°C | 80°C | 110°C |
| Твердость (по Роквеллу/Шору) | R105-R110 | R118 | R70-R90 | R70-R90 | R106 | D40 |
| Относительное удлинение при разрыве | 0% | 250% | 3,8% | 300% | 50% | 600% |
| Прочность на изгиб | 41 МПа | 36 МПа | 55,3 МПа | 70 МПа | 76,1 МПа | 5,3 МПа |
| Прочность на разрыв | 22 МПа | 34 МПа | 57,8 МПа | 66-83 МПа | 36,5 МПа | 17,5 МПа |
| Модуль упругости при растяжении | 1,6 ГПа | 1,35 ГПа | 3,3 ГПа | 2,7 ГПа | 2,6 ГПа | 0,86 ГПа |
| Модуль упругости при изгибе | 2,1 ГПа | 1,45 ГПа | 2,3 ГПа | 2,6 ГПа | 1,12 ГПа | 0,87 ГПа |
| Температура стеклования | 105°C | 95°C | 60-65°C | 50-70°C | 80°C | -10°C |
| Плотность | 1,1 г/см³ | 1,01 г/см³ | 1,23-1,25 г/см³ | 1,13 г/см³ | 1,3 г/см³ | 1,1 г/см³ |
| Температура экструзии | 210-245°C | 220-240°C | 190-230°C | 235-260°C | 215-245°C | 220-240°C |
| Температура стола | 90-120°C | 70-90°C | 20-60°C | 100-120°C | 20-80°C | 90-110°C |

Рис.1. Таблица параметров пластиков

Сначала описан словарь, который содержит диапазоны значений параметров для каждого типа пластика. На вход алгоритму подаются измеренные параметры материала. Программа создает словарь, где ключи – это типы пластика из словаря, а значения изначально равны нулю. Затем она проходит по всем типам пластика и их параметрам, сравнивая каждый параметр с введенным пользователем значением. Если значение параметра находится в пределах допустимого диапазона (с учетом погрешности в 10%), то для этого типа пластика увеличивается счетчик. После того как все параметры проверены, определяется максимальный балл и сумма всех баллов. Если сумма баллов равна нулю, выводится сообщение о том, что недостаточно данных для определения типа пластика. Если только один тип пластика имеет максимальный балл и этот балл равен количеству введенных параметров, то выводится название этого типа пластика. Если более одного типа пластика имеют максимальный балл, и этот балл равен количеству введенных параметров, или если максимальный балл меньше заданного процента (90% для учета погрешности измерений), выводится сообщение о недостаточности данных. Если не выполняется ни одно из условий выше, также выводится сообщение о недостаточности данных для определения типа

пластика. На рис.2 данный алгоритм представлен в виде блок-схемы.

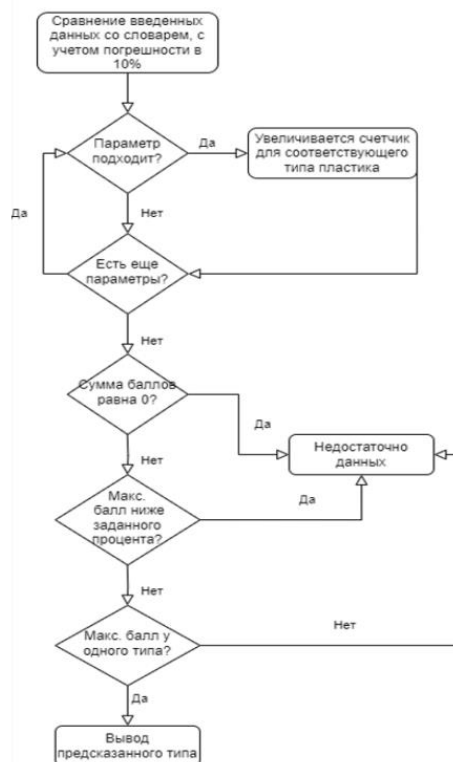


Рис.2. Алгоритм, представленный в виде блок-схемы

Чтобы сделать распознавание более устойчивым к ошибкам и погрешностям, была применена нейронная сеть – многослойный перцептрон [3], использующий полносвязные слои, с первым слоем в 50 скрытых нейронов и вторым слоем в 30 скрытых нейронов. Использовался фреймворк scikit-learn как простое и эффективное решение для задач машинного обучения. На основе исходной таблицы с параметрами и 10%-й погрешности был сгенерирован дата-сет измерений пластика, по 100 измерений на каждый класс. Все признаки были приведены к одному масштабу с помощью StandartScaler, и метки классов были закодированы с помощью LabelEncoder. Далее на подготовленных данных был обучен перцептрон [3]. Обученные компоненты сохраняются для дальнейшего получения предсказания. Данные вводятся аналогично алгоритму, и модель выводит предсказанный тип пластика.

Для решения задачи идентификации типа пластика по фотографии, нужно определить набор данных для обучения модели. Были взяты два наиболее распространенных типа – ABS и PLA. У пластика ABS структура более линейная (рис.3), а в PLA более пористая (рис.4).

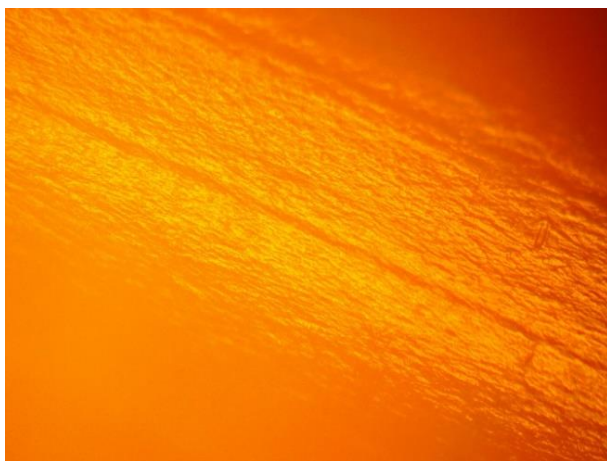


Рис.3. Пример структуры пластика ABS

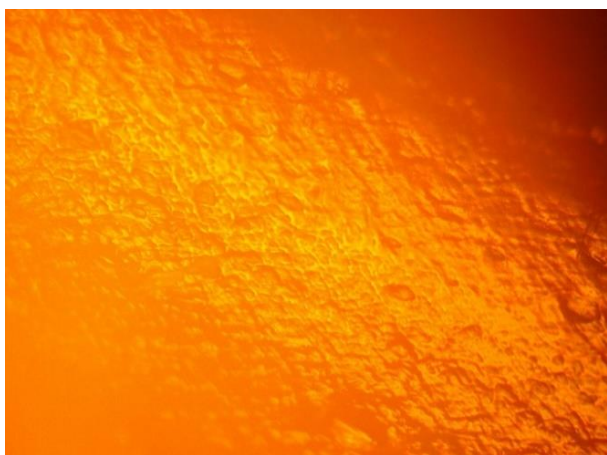


Рис.4. Пример структуры пластика PLA

Изображения этих пластиков также легче всего получить в достаточных для обучения нейросети количествах. Изображений иных типов либо единицы, либо их нет вовсе. Для проведения экспериментов был самостоятельно собран набор данных 66 изображений типа ABS и 72 изображений типа PLA. Все изображения размера 2592 на 1944, снятые на микроскоп в лаборатории на максимальном увеличении в 40 раз. Использовались материалы различных производителей, различных цветов, различной даты производства.

Фотографии преобразуются к размеру 224 на 224, для минимизации влияния цвета материала все изображения переводятся в оттенки серого. Для улучшения точности модели к тренировочным данным применяются различные аугментации данных, в виде поворота изображения, отражения по вертикали и горизонтали, добавления различных шумов. Размер тренировочной выборки составляет 80% всех изображений, валидационной 20%.

Используется фреймворк Tensorflow [1], являющийся одним из наиболее популярных и эффективных для задачи анализа изображений с помощью нейронных сетей, так как является гибким, имеет простой API и хорошо оптимизирован работе с GPU. Для создания нейронной сети использовалось дообучение нейронной сети MobileNetV3 [4] с добавлени-

ем двух полносвязных слоев. Данная модель предназначена для работающих в реальном времени приложений на мобильных устройствах и встраиваемых системах, что значительно расширяет возможность применения модели, и позволяет разместить систему на мобильных устройствах, ПЛИС или одноплатных компьютерах. Модель обеспечивает хорошую точность классификации изображений при малом объеме параметров и вычислительных ресурсов. Входными данными программы является путь к сохраненному на устройство изображению, выходными является строка с меткой предсказанного класса. Точность составила 91,25%.

В конце была предложена идея сегментировать изображение на 9 частей меньшего размера, чтобы не терять многие мелкие детали структуры пластика. Исходное изображение преобразуется в 9-мерный тензор вида (9, 864, 648, 1). Далее тензор передается в нейронную сеть с 3 сверточными [2] и 3 полносвязными слоями. К стандартным слоям применена оболочка TimeDistributed. Она предназначена для применения определенной операции к каждому шагу (кадру) входных данных независимо, что важно для учета структуры пластика и его различных деталей. Входными данными программы также является путь к сохраненному на устройство изображению, выходными также является строка с меткой предсказанного класса. Несмотря на перспективность идеи, достичь высокой точности не удалось, точность составила 64.29%.

Все 4 разработанные модели распознавания пластика объединены в единую оболочку и готовы к применению в API разработанной системы. Backend реализован с помощью библиотеки Flask [5] на языке Python. Такой язык выбран для максимальной совместимости с нейросетью, а библиотека из-за своей простоты и быстродействия. Процесс обработки запроса устроен следующим образом: на сервер поступает запрос, сервер обращается к нейросети и производится обработка, нейросеть отдаёт ответ серверу, сервер отдает ответ. Приложение имеет полную совместимость с Docker/Docker-compose. Проект имеет виртуальное окружение `pipenv`.

Frontend реализован с помощью библиотеки React на языке JavaScript. Библиотека выбрана из-за быстродействия и большой базы уже готовых и испытанных решений и компонентов. Для маршрутизации используется библиотека React-Router. Для отображения результатов обработки библиотека Toastify.

Также реализован телеграм-бот с помощью библиотеки `node-js-tg-bot-api` на языке JavaScript. Аналогично Frontend части сервиса бот имеет два варианта работы: Определение по параметрам. После ввода команды `/recognize_image_by_params` и ввода параметров бот выдаёт результат. Определение по фотографии. После ввода команды `/recognize_image_by_photo` и отправки фотографии бот выдаёт результат. Схематично работа данных сервисов представлена на рис.5.

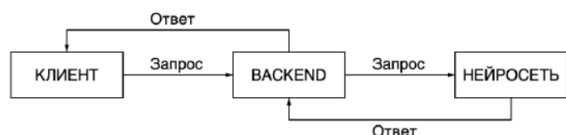


Рис.5. Описание работы сервиса

Аппаратная часть решения

Для передачи данных в программную часть, параметры пластика необходимо измерить. Необходимо спроектировать установку способную определять свойства пластика при растяжении. Также требуется установка для автоматического фотографирования пластика под микроскопом.

Разрывное устройство (рис.6) предназначено для закрепления в нем пластика и дальнейшего его разрыва, контролируя скорость и силу растяжения, для определения механических характеристик пластика. Модель устройства выполнена в виде одного сборочного объекта. Материалом изготовления корпуса была выбрана листовая сталь из-за прочностных характеристик, дешевизны и простоты обработки. Разрывная установка выполнена в виде параллелепипеда без стенок. В качестве силовых элементов используются профильные уголки толщиной 5 мм которые крепятся к крепежной пластине двигателя такой же толщиной на усиленные крепежные уголки. Все детали скрепляются друг с другом болтами, что упрощает сборку. Механизм растяжения выполнен из 2х двигателей постоянного тока Dynamo sliven pik 12-3, к которым через муфты подсоединены 2 шпильки, расположенные вертикально. На шпильках расположен бегунок – крепление верхней лапки которое имеет нарезанную резьбу для перемещения вдоль шпилек. Нижняя лапка крепится к тензодатчику, который в свою очередь через переходник крепится к пластине 4-х точечным креплением. Снизу на двигателях установлены инкрементальные энкодеры позволяющие точно удерживать необходимую угловую скорость моторов при увеличивающейся на них нагрузке.

Фотографирующее устройство (рис.6) предназначено для установки перед экструдером 3d принтера для автоматического определения типа пластика путем его фотографирования. Модель устройства выполнена из ABS пластика т.к. имеет сложные элементы и нагрузка на него будет небольшой. Из стали выполнены шпильки, гайки на них находящиеся и крепление объектива. Задача данного устройства класть пластик перед объективом камеры ровно и на одной высоте, чтобы не сбился фокус на камере. По этой причине жёлоба в нижней части устройства узкие, чтобы на ровном участке нить распрямилась, тянуть её будет сам 3d принтер.

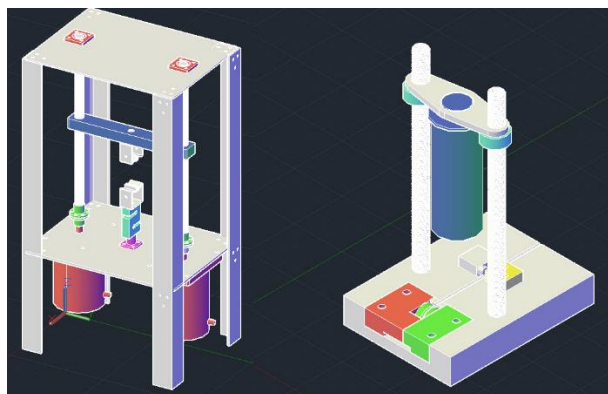


Рис.6. Разрывная и фотографирующая установки

Заключение

Был реализован программно-аппаратный комплекс для определения типа пластика для 3-Д печати, включающий в себя алгоритм для идентификации типа по измеренным параметрам, нейронную сеть для определения типа по фотографии, а также две установки, одну для определения параметров, другую для фотографирования материала на микроскоп. Был проанализирован рынок полимерных материалов для 3Д печати, выявлены возможности и методы для определения их параметров, а также созданы довольно большие массивы данных. Достигнута высокая точность работы нейронной сети и алгоритма, а также продемонстрирована практическая возможность создания данного комплекса из существующих на рынке методов и технологий. В качестве долгосрочных вариантов развития проекта можно рассмотреть вариант создания открытого API для автоматического доступа к сервисам из крупных центров 3Д печати, которые обладают необходимым оборудованием для измерения, а также разработка и реализация собственных измерительных установок. Также в дальнейшем можно добавить в установку больше возможностей для измерения различных параметров, а в нейронную сеть и алгоритм добавить больше разновидностей пластика.

Список литературы

1. Библиотека tensorflow – URL: <https://www.tensorflow.org/?hl=ru> (дата обращения: 10.01.2024). – Текст : электронный.
2. Сверточные нейронные сети – URL: <https://www.geeksforgeeks.org/introduction-convolution-neural-network/> (дата обращения: 14.01.2024). – Текст : электронный.
3. Многослойный перцептрон – URL: <https://www.geeksforgeeks.org/what-is-perceptron-the-simplest-artificial-neural-network/> (дата обращения: 15.01.2024). – Текст : электронный.
4. Нейронная сеть MobileNetV3 – URL: <https://arxiv.org/abs/1905.02244> (дата обращения: 17.01.2024). – Текст : электронный.
5. Голубничая, Я. Р., Проскураков Н. Е. Перспективные технологии 3Д-печати – URL: <https://www.geeksforgeeks.org/flask-creating-first-simple-application/> (дата обращения: 24.01.2024). – Текст : электронный.

ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕВОГО АНАЛИЗА ЦИФРОВЫХ РЕНТГЕНОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ ОРГАНОВ ГРУДНОЙ КЛЕТКИ (ФЛЮОРОГРАФИИ) ДЛЯ ДИАГНОСТИКИ ПАТОЛОГИЙ ЛЕГКИХ С ПОМОЩЬЮ МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Кочарян Д.Р.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ*

Аннотация

Пневмония является распространенной болезнью, которая может привести к летальному исходу. В ее лечении очень важно своевременно поставить диагноз, в этом случае могут помочь нейронные сети анализирующие рентгенограммы. Этот процесс можно автоматизировать с помощью методов компьютерного зрения. На данный момент есть множество статей, в которых описаны модели компьютерного зрения для решения этой задачи. Однако большинство из них использовало для обучения менее 2000 снимков, что не гарантирует адекватность результата. В данном исследовании будут сравниваться хорошо показавшие себя модели IVGG13 и DarkCovidNet, для которых будет использовано более 10000 рентгенограмм. В результате исследования обе модели показали приемлемый результат, однако DarkCovidNet превзошла по всем метрикам.

Введение

Пневмония, осложнение вирусных инфекций вроде COVID-19 или гриппа, согласно Всемирной организации здравоохранения, ежегодно уносят жизни до 50 тысяч человек, в то время как число активных случаев COVID-19 достигает 8,2 миллиона, с числом смертей 445 535. Анализ рентгеновских снимков является средством для выявления и ранней диагностики заболевания и дает быстрые и точные результаты, что делает лечение более эффективным [1], [2].

Современные методы автоматического анализа рентгенограмм позволяют упростить работу радиологов в чтении снимка и постановке диагноза[3]. Однако они не способны заменить человека и рассматриваются только в качестве вспомогательного инструмента[4].

Такие методы как правило основаны на использовании глубоких нейронных сетей [5]–[7]. Ключевую роль для моделей играет датасет, на котором они обучались, чем больше объем, тем больше зависимостей получается выучить. В данной работе использовался набор данных состоящий из 5856 рентгенограмм.

В этой статье будут рассматриваться две нейросетевые модели. Первая модель от Zhi-Peng Jiang из статьи [8] её название IVGG13. Для ее обучения использовалось порядка 5 000 рентгенограмм, расширенных с помощью аугментации. Показатели

метрик Accuracy:77.5; Precision: 73.5; Recall: 99.5; F1-score: 84.6. Вторая модель сделана Tulin Ozturk на основе DarkNet ее название DarkCovidNet[9]. Обучалась на 1000 снимков. Ее результаты составили Accuracy: 98.08; Precision: 98.03; Recall: 95.3;F1-score: 95.51. Обе сети показали хорошие результаты, но на различных данных, что не позволяет их сравнить.

Целью данного исследования является сравнение эффективности двух моделей, обученных на одной выборке.

Сбор и подготовка набора данных

В данной работе используется набор из 5856 рентген-снимков грудной клетки, разделенных 4 к 1 для обучения и теста. Пример изображения приведен на рис.1. Для расширения выборки использовались такие методы аугментации данных, как горизонтальное отражение и поворот на случайный угол до 30 градусов в обе стороны от часовой стрелки. Данные преобразования позволили увеличить число изображений минимум в 2 раза. После аугментации снимки приводились к размеру 256x256x3 для DarkCovidNet и 128x128x3 для IVGG13, затем производится нормализация с параметрами математического ожидания 0.5 и дисперсии 0.5 для всех каналов. На следующем шаге данные разбиваются на пакеты, содержащие по 64 изображения, и подаются на обучение модели.

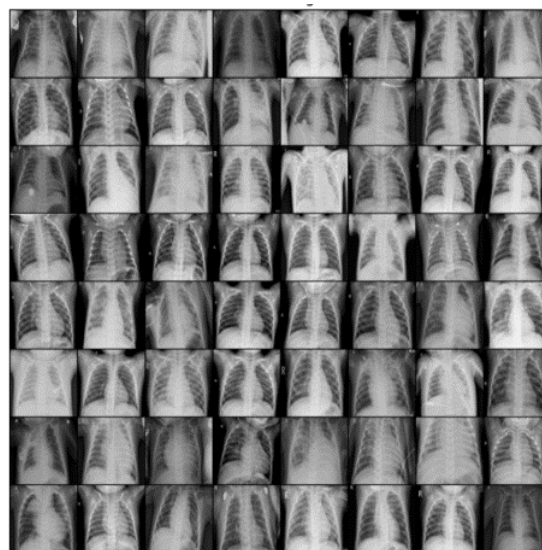


Рис.1. Пример данных

Построение и обучение моделей

Модели были построены на языке python с помощью библиотеки pytorch, их структура была оформлена согласно статьям [8] для IVGG13 (Таблица 1. Структура IVGG13) и [9] для DarkCovidNet (Таблица 2 Структура сети DarkCovidNet). В качестве функции ошибки использовалась кросс-энтропийная, оптимизатором был Adam для обеих моделей. Обучение производилось до тех пор, пока метрики Accuracy, Precision, Recall, F1-score не вышли на плато.

Таблица 1. Структура IVGG13

| Layer type | Output shape |
|--------------------------------|----------------|
| conv2d_1 (conv2D) | [128, 128, 32] |
| conv2d_2 (conv2D) | [128, 128, 32] |
| max_Pooling2d_1 (Max-Pooling2) | [64, 64, 32] |
| conv2d_3 (conv2D) | [64, 64, 32] |
| conv2d_4 (conv2D) | [64, 64, 32] |
| max_Pooling2d_2 (MaxPooling2) | [32, 32, 32] |
| conv2d_5 (conv2D) | [32, 32, 64] |
| conv2d_6 (conv2D) | [32, 32, 64] |
| max_Pooling2d_3 (MaxPooling2) | [16, 16, 64] |
| conv2d_7 (conv2D) | [16, 16, 128] |
| conv2d_8 (conv2D) | [16, 16, 128] |
| max_Pooling2d_4 (MaxPooling2) | [8, 8, 128] |
| conv2d_9 (conv2D) | [8, 8, 64] |
| conv2d_10 (conv2D) | [8, 8, 64] |
| max_Pooling2d_5 (MaxPooling2) | [4, 4, 64] |
| flatten_1 (Flatten) | [1024] |
| dense_1 (Dense) | [1024] |
| dropout_1 (Dropout) | [1024] |
| dense_2 (Dense) | [1024] |
| dropout_2 (Dropout) | [1024] |
| dense_3 (Dense) | [2] |

Таблица 2. Структура DarkCovidNet

| Layer type | Output shape |
|------------|----------------|
| Conv2d | [8, 256, 256] |
| Conv2d | [16, 128, 128] |
| Conv2d | [32, 64, 64] |
| Conv2d | [16, 66, 66] |
| Conv2d | [32, 66, 66] |
| Conv2d | [64, 33, 33] |
| Conv2d | [32, 35, 35] |
| Conv2d | [64, 35, 35] |
| Conv2d | [128, 17, 17] |
| Conv2d | [64, 19, 19] |
| Conv2d | [128, 19, 19] |
| Conv2d | [256, 9, 9] |
| Conv2d | [128, 11, 11] |
| Conv2d | [256, 11, 11] |
| Conv2d | [128, 13, 13] |
| Conv2d | [256, 13, 13] |
| Conv2d | [2, 13, 13] |
| Flatten | [338] |
| Linear | [2] |

Вследствие обучения моделей на тренировочной выборке были получены значения кросс-

энтропийной функции ошибки равные 0.0055 и 0.0034 для IVGG13 и DarkCovidNet соответственно. Значения метрик на тестовой выборке для первой сети Accuracy: 0.7949; Precision:0.8325; Recall: 0.7493; F1-score: 0.7736, для второй модели Accuracy: 0.8729; Precision:0.8952; Recall:0.8294; F1-score: 0.8610 (см. Таблица 3) . В случае с IVGG13 результаты схожи с теми, что описывал автор в статье [8]. Исключение составляют Recall и F1-score, разница равна 25.93 и 7.24. Однако показатели DarkCovidNet упали, в оригинале все метрики больше 0.95, когда в реплики наибольшее значение имеет Precision равный 0.8952. Разница между моделями из статей [8], [9] и построенными в этой работе объясняется различными данными и гиперпараметрами. По окончании обучения нейросетей DarkCovidNet имеет по всем метрикам большее качество относительно IVGG13.

Таблица 3. Значения метрик качества

| Название модели | Accuracy | Precision | Recall | F1-score |
|-----------------|----------|-----------|--------|----------|
| DarkCovid Net | 0.7949 | 0.8325 | 0.7493 | 0.7736 |
| IVGG13 | 0.8729 | 0.8952 | 0.8924 | 0.8294 |

Заключение

В результате данной работы были обучены модели DarkCovidNet и IVGG13 на одном и том же наборе данных, что обеспечило возможность их сравнения. В итоге модель DarkCovidNet превзошла по всем метрикам конкурента. Обе нейросети нельзя использовать без участия человека, так как показатели слишком низкие, однако они хорошо проявят себя в роли вспомогательного инструмента для радиологов в задаче диагностики пневмонии, что в будущем может быть использовано для автоматизации процесса формирования диагноза заболевания.

Список литературы

1. G. Xu *et al.*, "Clinical Pathway for Early Diagnosis of COVID-19: Updates from Experience to Evidence-Based Practice," *Clin. Rev. Allergy Immunol.*, vol. 59, no. 1, pp. 89–100, 2020, doi: 10.1007/s12016-020-08792-8.
2. D. Dansana *et al.*, "Early diagnosis of COVID-19-affected patients based on X-ray and computed tomography images using deep learning algorithm," *Soft Comput.*, vol. 27, no. 5, pp. 2635–2643, 2023, doi: 10.1007/s00500-020-05275-y.
3. F. Caobelli, "Artificial intelligence in medical imaging: Game over for radiologists?," *Eur. J. Radiol.*, vol. 126, no. February, p. 108940, 2020, doi: 10.1016/j.ejrad.2020.108940.
4. D. Jin *et al.*, "Artificial intelligence in radiology," *Artif. Intell. Med. Tech. Basis Clin. Appl.*, pp. 265–289, 2020, doi: 10.1016/B978-0-12-821259-2.00014-4.
5. P. Kumar Sethy, S. Kumari Behera, P. Kumar Ratha, and P. Biswas, "Detection of coronavirus Disease

(COVID-19) based on Deep Features and Support Vector Machine,” 2020. [Online]. Available: www.preprints.org

6. Y. Song *et al.*, “Deep Learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT Images,” *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 18, no. 6, pp. 2775–2780, 2021, doi: 10.1109/TCBB.2021.3065361.

7. D. Khemasuwan, J. S. Sorensen, and H. G. Colt, “Artificial intelligence in pulmonary medicine: Computer vision, predictive model and covid-19,” *Eur. Respir. Rev.*, vol. 29, no. 157, pp. 1–16, 2020, doi: 10.1183/16000617.0181-2020.

8. Z. P. Jiang, Y. Y. Liu, Z. E. Shao, and K. W. Huang, “An improved VGG16 model for pneumonia image classification,” *Appl. Sci.*, vol. 11, no. 23, 2021, doi: 10.3390/app112311185.

9. T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, “Automated detection of COVID-19 cases using deep neural networks with X-ray images,” *Comput. Biol. Med.*, vol. 121, Jun. 2020, doi: 10.1016/j.compbiomed.2020.103792.

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ДОБАВЛЕНИЯ ИНТЕРАКТИВНЫХ ЭЛЕМЕНТОВ В ВИРТУАЛЬНЫЕ ПРОСТРАНСТВА ДЛЯ ПРОВЕДЕНИЯ ВСТРЕЧ ИЛИ КОРПОРАТИВНЫХ МЕРОПРИЯТИЙ И ИХ ВЛИЯНИЕ НА ПОЛЬЗОВАТЕЛЕЙ

Чуйко Д.О., Кретова А.А.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Данная научно-техническая работа фокусируется на исследовании возможностей интеграции интерактивных элементов в виртуальные пространства для эффективного улучшения проведения встреч и корпоративных мероприятий. Авторы представляют обзор современных технологий, оценивают их технические возможности, а также анализируют влияние интерактивности на взаимодействие пользователей, создавая базу для повышения вовлеченности и эффективности виртуальных встреч.

Введение

С развитием виртуальных технологий сфера корпоративных встреч и мероприятий переживает существенные изменения, акцентируя внимание на виртуальных пространствах. В данном контексте важным направлением становится интеграция интерактивных элементов, создающих более насыщенный и эффективный опыт для участников. Цель данного исследования заключается в изучении возможностей добавления интерактивных элементов в виртуальные пространства с целью улучшения проведения корпоративных встреч и мероприятий.

В задачи исследования входит проведение обзора литературы и анализа существующих технологий в данной области, рассмотрение и разработка механизма реализации технических аспектов внедрения интерактивности, анализ влияния таких элементов на взаимодействие участников.

В первом разделе исследования представлен обзор литературы и существующих технологий, а также акцентированы ключевые тренды в области виртуальных пространств для корпоративных мероприятий. Второй раздел анализирует технические аспекты внедрения интерактивности. Что касается третьего раздела, то он фокусируется на влиянии этих элементов на взаимодействие участников. В заключении представлены основные выводы и рекомендации, основанные на результатах проведенного исследования.

Обзор литературы, существующие технологии и ключевые тренды в области разработки виртуальных пространств

В современном мире виртуальные пространства становятся неотъемлемой частью корпоративных мероприятий, предоставляя участникам возможность взаимодействия, не зависящего от их физического местоположения. Обзор литературы в данной области выявляет значительный интерес к применению технологий виртуальной и дополненной реальности.

Большой успех в недавнее время был достигнут в секторе виртуального образования. Метавселенная разрешила многие трудности в этой области. Особенно в ситуации с COVID-19 люди продолжали все образовательные мероприятия виртуально, используя технологии метавселенной. Это был один из малого количества секторов, развитию которого эпидемия не мешала, а только способствовала. [1]

Метавселенная, по мнению одной из глобальных корпораций, должна предоставить широкие возможности онлайн-пользователям не только в сфере коммуникации, проведения досуга, но и обучения, организации своего бизнеса и т. д., а непосредственно глобальный цифровой мир должен стать максимально реалистичным. [2]

Существует ряд ключевых трендов, которые формируют современное виртуальное пространство для корпоративных мероприятий. Наблюдается рост интереса к интеграции технологий виртуальной и дополненной реальности для создания более реалистичных и вовлекающих мероприятий. [3] Основными трендами также являются активное использование виртуальных платформ для проведения совещаний и тимбилдинга, что позволяет участникам более глубоко взаимодействовать в виртуальной среде. [4]

Интерактивность играет важную роль в обеспечении эффективного взаимодействия пользователей виртуальных пространств. На примере разработки интерактивного виртуального учебного класса говорится о таких преимуществах как повышенная вовлеченность учащихся и активная совместная работа. Также подчеркивается важность продолжения разработок и исследований для полной реализации потенциала метавселенной в революционизировании образования. [5]

Эти научные работы формируют основу для понимания текущего состояния и тенденций в области виртуальных пространств для корпоративных мероприятий и подчеркивают роль интерактивных элементов в улучшении виртуального взаимодействия.

Этапы разработки виртуального пространства с помощью Unity

Для разработки виртуальных пространств выбрана среда Unity - мощный игровой движок, поддерживающий широкий спектр платформ. Его преимущества включают в себя высокий уровень графической обработки, многоплатформенность, гибкость программирования на языке C#, а также активное сообщество и обширная ресурсная база Asset Store для удобства разработки.

Этапы разработки виртуального пространства с использованием Unity включают следующие шаги: создание основной структуры сцены, разработка 3D моделей объектов и их расположение на сцене, добавление интерактивных элементов, реализация отдельных функций, к примеру, для удобного перемещения между различными локациями, а также тестирование и оптимизация виртуального пространства перед его внедрением. Эти этапы обеспечивают создание качественного и функционального виртуального пространства, способного эффективно взаимодействовать с пользователем и обеспечивать ему приятный и полезный опыт. [6,7]

Технические возможности и реализация интерактивных элементов при разработке виртуальных пространств

Благодаря гибкости среды разработки возможно реализовать интерактивные элементы под любые потребности. Для проведения встреч и корпоративных мероприятий наиболее полезными могут служить следующие элементы.

1. Интерактивные кнопки могут служить для запуска определенных действий, открытия меню, или активации функциональности виртуального пространства. В Unity создаются 3D-модели кнопок и присваиваются им триггеры для обработки событий нажатия. Сценарии на языке программирования C# реализуют логику действий при нажатии.

2. Перемещаемые объекты. Позволяет пользователям взаимодействовать с объектами, перемещая их в виртуальном пространстве, что может быть полезно для организации окружения или решения головоломок. Объектам в Unity присваиваются физические свойства, и программные функции обрабатывают события мыши для захвата и перемещения объектов.

3. Создание виртуальных мест для сидения может быть полезным в виртуальных мирах, где пользователи могут взаимодействовать с окружением, предоставляя им уникальные точки обзора. Также это может быть полезно при проведении конференций, ежедневных встреч или стратегических сессий, где важно, чтобы персонажи не толпились в одном месте, а были грамотно рассажены по переговорной комнате. В Unity создаются коллайдеры или зоны, и обработка событий определяет вход пользователя в

зону, что может вызывать изменение состояния на сидение. Есть возможность добавления специальных интерактивных элементов, при нажатии на которые персонаж пользователя может сесть (рис.1).



Рис.1. Интерактивный элемент, позволяющий сесть на кресло в переговорной

Реализация функции, отвечающей за присаживание отображена на рис.2.

```
void SitDown()
{
    // Логика присаживания на стул
    isSitting = true;
    // Отключение движения пользователей
    // во время сидения на стуле
    characterController.enabled = false;
    transform.position = sittingPosition.position;
}
```

Рис.2. Фрагмент кода, отвечающий за логику присаживания на стул

4. Телепорты предоставляют средство перемещения пользователя в разные части виртуального пространства, улучшая мобильность и удобство навигации. Также это позволяет перемещаться между различными виртуальными мирами. Внутри каждого мира есть ограничения на количество ресурсов, используемых в окружении, для наиболее быстрой работы. Чтобы обойти эти ограничения, рекомендуется разные локации или комнаты одного пространства располагать в разных мирах.

Реализуется так, что в Unity создаются точки телепорта и обработчики событий, которые активируют телепортацию при взаимодействии пользователя с соответствующими элементами. Встроенные функции Unity обеспечивают процесс телепортации.

5. Экраны для презентаций предоставляют возможность виртуальных лекций, просмотра мультимедийного контента или проведения виртуальных митапов. Создаются 3D-модели экранов с использованием видеоплеера Unity для воспроизведения презентаций, а затем с помощью программных функций возможно управление событиями воспроизведения.

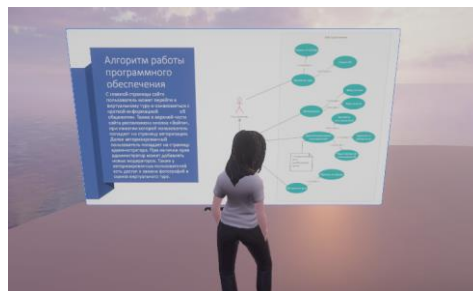


Рис.3. Вывод на экран презентации

Влияние интерактивных элементов на пользователей виртуальных площадок

В ходе научного эксперимента были разработаны два виртуальных мира с целью сравнительного анализа воздействия интерактивных элементов на восприятие и удобство использования виртуальных пространств. Первый мир был обогащен интерактивными элементами, такими как возможность сидеть на стульях, открывать двери, а также использовать телепортацию для быстрого перемещения между различными локациями. Второй мир представлял собой простую трехмерную сцену без интерактивности.

Проведенный эксперимент показал, что виртуальное пространство с интерактивными элементами оказалось более удобным и функциональным для пользователей. В таблице 1 приведены результаты опроса участников эксперимента, в каком виртуальном мире они бы предпочли проводить встречи. Участники эксперимента высоко оценили возможность взаимодействия с окружающим миром, что сделало виртуальное пространство более привлекательным и удобным для использования. В частности, участники отметили улучшение восприятия информации, повышение удовлетворенности от процесса взаимодействия и увеличение общей эффективности использования виртуальной среды. Таким образом, результаты эксперимента подтвердили положительное влияние интерактивных элементов на пользовательский опыт и подчеркнули их важность для создания более функциональных и привлекательных виртуальных пространств.

Таблица 4. Процент участников, которые предпочли бы проведение встреч в определенном мире

| Простое виртуальное пространство | Виртуальное пространство с интерактивными элементами |
|----------------------------------|--|
| 8% | 92% |

Заключение

В заключении данной научной работы подчеркивается значимость интерактивных элементов в виртуальных пространствах при организации встреч и корпоративных мероприятий. Эти элементы существенно влияют на восприятие пользователей, предоставляя участникам более глубокий и вовлекающий опыт взаимодействия в метавселенных. Внедрение интерактивных кнопок, виртуальных презентационных экранов, а также совместных пространств для обсуждений приводит к созданию динамичных и эффективных виртуальных сценариев для встреч и корпоративных событий. Эти решения не только оптимизируют организацию событий, но и способствуют формированию более тесного социального взаимодействия между участниками.

С теоретической точки зрения работа вносит вклад в область исследований виртуальных пространств, подчеркивая важность интерактивности для улучшения пользовательского опыта. Анализ теоретических подходов к использованию интерактивных элементов в контексте виртуальных встреч

обогащает теоретическую базу знаний, обеспечивая более глубокое понимание влияния этих элементов на восприятие и вовлеченность участников.

Практическая значимость настоящей работы заключается в предоставлении конкретных рекомендаций и практических решений для внедрения интерактивных элементов в виртуальные пространства при организации встреч и корпоративных мероприятий. Исследование выделяет потенциал этих элементов для оптимизации виртуальных событий, улучшения взаимодействия участников и создания более продуктивной рабочей атмосферы.

Список литературы

1. Islam Mozumder, M. A. Technological Roadmap of the Future Trend of Metaverse based on IoT, Blockchain, and AI Techniques in Metaverse Education / M. A. Islam Mozumder, A. Athar, A. Poupri, M. M. Sheeraz. - Текст: электронный // 25th International Conference on Advanced Communication Technology (ICACT). - 2023.- С. 1414-1423. - URL: https://www.researchgate.net/publication/370960266_Technological_Roadmap_of_the_Future_Trend_of_Metaverse_e_based_on_IoT_Blockchain_and_AI_Techniques_in_Metaverse_Education (дата обращения: 25.01.2024).
2. Ваторопин, А. С. Метавселенная: перспективны создания и социальные последствия / А. С. Ваторопин, С. А. Ваторопин, И. И. Тепляков - Текст: электронный // Теория и практика общественного развития. - 2023.- №4 (170). - URL: <https://dom-hors.ru/teoriapraktika/2022/4> (дата обращения: 27.01.2024).
3. Florea, C. Extending a User Involvement Tool with Virtual and Augmented Reality / C. Florea, P. Alavesa, L. Arhippainen. - Текст: электронный // IEEE Conference on Virtual Reality and 3D User Interfaces (VR) - 2019.- С. 925-926. - URL: <https://ieeexplore.ieee.org/document/8798299> (дата обращения: 27.01.2024).
4. Филипова, И. А. Метавселенные: как их развитие повлияет на работников и работодателей / И. А. Филипова. - Текст: электронный // Ежегодник трудового права - 2019.- №13. - С. 45-64. - URL: <https://rjll.spbu.ru/article/download/13497/10355/> (дата обращения: 29.01.2024).
5. Seiri S. AI Exploring Immersive Learning in the Metaverse: A Prototype for Interactive Virtual Classroom / S. Al Seiri, H. Al Kaabi and J. N. Al-Karaki. - Текст: электронный // International Conference on Intelligent Metaverse Technologies & Applications (iMETA) - 2023. - С. 1-8. - URL: <https://ieeexplore.ieee.org/document/10294515> (дата обращения: 29.01.2024).
6. Dixit A. Creating metaverse through Unity 3D and Blender / A. Dixit - Текст: электронный // ResearchGate - 2023. - С. 1-3. - URL: https://www.researchgate.net/publication/370751027_CREATING_METAVERSE_THROUGH_UNITY_3D_AND_BLENDER (дата обращения: 01.02.2024).
7. Unity : Road to the Metaverse [сайт]. – 2023. – URL: <https://unity.com/ru/resources/workshop-12-build-geolocated-and-social-experiences-in-ar?elqTrackId=135868ee5eb74ca38308c381ff4415ed&elqaid=4090&elqat=2> (дата обращения 24.01.2024)

РАЗРАБОТКА МЕТОДОЛОГИИ ОЦЕНИВАНИЯ СТЕПЕНИ СУГГЕСТИВНОСТИ РУССКОЯЗЫЧНЫХ РЕКЛАМНЫХ ТЕКСТОВ

Блюдова В.М.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент фундаментальной и
прикладной лингвистики*

Аннотация

Реклама – неотъемлемая часть рынка. Цель рекламы – убедить потребителя в необходимости приобретения рекламируемого продукта. Данная цель достигается посредством создания рекламных текстов с наибольшим суггестивным потенциалом. Возникает необходимость численной оценки суггестивности. В данной работе предлагается подход, который позволит решить данную задачу, и демонстрируется работа предложенной методологии на примере.

Введение

На тему изучения влияния рекламы на сознание человека опубликовано немало исследований. Большинство работ направлены на разработку классификации суггестии или на извлечение суггестивных средств. Однако для копирайтеров определить, какие способы внушения использовать, зачастую недостаточно, необходимо понимать, насколько текст суггестивен. Мы обнаружили, что в научном поле отсутствуют работы по оценке суггестивности, и предложили подход для решения проблемы. Таким образом, целью данного исследования – разработать методологию для определения внушаемости рекламного текста.

Теоретические основы

В данном исследовании, вслед за Сидоренко Е.В. [8], мы трактуем термин суггестия (суггестивности), как «сознательное неаргументированное воздействие на человека или группу людей, имеющее своей целью изменение их состояния, отношения к чему-либо и создания предрасположенности к определенным действиям». Также необходимо ввести понятие суггестивности.

В рамках данной работы мы проводим анализ рекламных текстов на лексическом и грамматическом уровнях языка. За основу была взята классификация Цукановой Е.М. [9]. В работе "Языковые средства выражения суггестии в рекламном слогане" автор выделяет десять лексических и грамматических средств. Однако данная классификация была составлена Е.М.Цукановой [9] на основе материалов на немецком языке. По этой причине перед нами стояла задача убедиться в том, что в русском языке данные средства также используются для оказания воздействия. Мы обратились к работам Селемёновой О.А. [7], Бернадской Ю.С. [4], Бойчук Е.И. [5], Голайденко Л.Н. [6] и Шелестюк Е.В. [10], в которых авторы пишут о суггестивных средствах, используемых в

русскоязычном дискурсе. Рассмотрим данные признаки подробнее.

1. Имена собственные. В обществе существуют некоторые сложившиеся стереотипные представления, связанные с определёнными продуктами потребления, произведенными в те или иные города и странах.

2. Слова с конкретным значением. Использование конкретных слов способствует формированию более точного представления потребителя о рекламируемом товаре.

3. Заимствованная лексика. Данное средство способствует формированию эффекта доверия за счет восприятия носителем языка заимствованной лексики как более престижной, актуальной и привлекательной.

4. Лексические повторы, которые позволяют создать ритмичность в тексте. В свою очередь ритмичность является одним из просодических средств внушения, по словам Бойчук Е.И. [5].

5. Антонимы придают суггестивность тексту за счет парадоксальности.

6. Числительные способствуют положительной оценке предмета или услуги и выступают средством реализации «стратегии преимущества».

7. Устойчивые выражения. В данной работе под устойчивыми выражениями мы понимаем фразеологизмы, которые используются, как правило, для создания комического эффекта, обусловленного несоответствием данного социального явления объективному ходу вещей, что ведет к запоминаемости.

8. Повелительное наклонение глаголов побуждает покупателя к совершению действия.

9. Прилагательные в сравнительной и превосходной степени обладают яркими экспрессивными свойствами, формируют положительную оценку объекта рекламы.

10. Пассивный залог. Шелестюк Е.В. [10] пишет о том, что указание конкретных лиц может включить психологическую защиту тех, кто не разделяет высказанное мнение, что затрудняет эффект внушения, поэтому в рекламе зачастую используют прием пассивизации.

Методология для оценки степени суггестивности текста

В рамках данного исследования мы предлагаем методологию, которая позволяет определить в процентах суггестивность текста. Работа состояла из следующих этапов:

Этап 1. Создание опроса для получения оценок степени суггестивности признаков.

Для того, чтобы определить степень суггестивности рекламного текста, необходимо выяснить, какой вес каждый отдельный признак вкладывает в общую суггестивность. В данной работе для получения весов признаков мы применяем метод анализа иерархий Саати Т.Л. [2].

В качестве критерия для сравнения выступает суггестивность, в качестве альтернатив мы используем 10 признаков суггестивности, которые обозна-

чаются, как F1 – F10. Для шкалы оценивания были отобраны следующие значения (Таблица 1).

Таблица 1. Шкала оценивания степени суггестивности признака

| | |
|---------|---|
| 1 | признаки F1 и F2 одинаково суггестивны |
| 3 (1/3) | признак F1 чуть более (менее) суггестивный, чем признак F2 |
| 5 (1/5) | признак F1 более (менее) суггестивный, чем признак F2 |
| 7 (1/7) | признак F1 значительно более (менее) суггестивный, чем признак F2 |
| 9 (1/9) | признак F1 принципиально более (менее) суггестивный, чем признак F2 |

Этап 2. Оценивание степени суггестивности признаков.

В качестве экспертов были приглашены два кандидата филологических наук и магистр в области лингвистики.

В результате проведения опроса мы получили следующие матрицы экспертных оценок.

Таблица 2. Результат экспертных оценок Эксперта 1

| | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 |
|-----|------|------|------|------|------|------|------|------|------|------|
| F1 | 1,00 | 9,00 | 1,00 | 1,00 | 9,00 | 1,00 | 9,00 | 0,33 | 7,00 | 1,00 |
| F2 | 0,11 | 1,00 | 0,11 | 0,11 | 1,00 | 0,11 | 1,00 | 0,11 | 0,11 | 0,11 |
| F3 | 1,00 | 9,00 | 1,00 | 1,00 | 9,00 | 9,00 | 9,00 | 3,00 | 7,00 | 1,00 |
| F4 | 1,00 | 9,00 | 1,00 | 1,00 | 9,00 | 1,00 | 9,00 | 1,00 | 7,00 | 1,00 |
| F5 | 0,11 | 1,00 | 0,11 | 0,11 | 1,00 | 0,11 | 1,00 | 0,11 | 0,11 | 0,11 |
| F6 | 1,00 | 9,00 | 0,11 | 1,00 | 9,00 | 1,00 | 9,00 | 1,00 | 7,00 | 1,00 |
| F7 | 0,11 | 1,00 | 0,11 | 0,11 | 1,00 | 0,11 | 1,00 | 0,11 | 0,11 | 0,11 |
| F8 | 3,00 | 9,00 | 0,33 | 1,00 | 9,00 | 1,00 | 9,00 | 1,00 | 1,00 | 1,00 |
| F9 | 0,14 | 9,00 | 0,14 | 0,14 | 9,00 | 0,14 | 9,00 | 1,00 | 1,00 | 0,14 |
| F10 | 1,00 | 9,00 | 1,00 | 1,00 | 9,00 | 1,00 | 9,00 | 1,00 | 7,00 | 1,00 |

Таблица 3. Результат экспертных оценок Эксперта 2

| | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 |
|-----|------|------|------|------|------|------|------|------|------|------|
| F1 | 1,00 | 5,00 | 1,00 | 5,00 | 5,00 | 1,00 | 3,00 | 1,00 | 5,00 | 3,00 |
| F2 | 0,20 | 1,00 | 0,14 | 0,20 | 0,33 | 0,14 | 0,20 | 0,14 | 0,33 | 0,33 |
| F3 | 1,00 | 7,00 | 1,00 | 5,00 | 9,00 | 1,00 | 7,00 | 1,00 | 7,00 | 7,00 |
| F4 | 0,20 | 5,00 | 0,20 | 1,00 | 5,00 | 0,33 | 0,33 | 0,33 | 0,20 | 0,20 |
| F5 | 0,20 | 3,00 | 0,11 | 0,20 | 1,00 | 0,11 | 0,14 | 0,11 | 0,33 | 0,33 |
| F6 | 1,00 | 7,00 | 1,00 | 3,00 | 9,00 | 1,00 | 5,00 | 1,00 | 5,00 | 5,00 |
| F7 | 0,33 | 5,00 | 0,14 | 3,00 | 7,00 | 0,20 | 1,00 | 0,33 | 0,33 | 0,20 |
| F8 | 1,00 | 7,00 | 1,00 | 3,00 | 9,00 | 1,00 | 3,00 | 1,00 | 7,00 | 7,00 |
| F9 | 0,20 | 3,00 | 0,14 | 5,00 | 3,00 | 0,20 | 3,00 | 0,14 | 1,00 | 1,00 |
| F10 | 0,33 | 3,00 | 0,14 | 5,00 | 3,00 | 0,20 | 5,00 | 0,14 | 1,00 | 1,00 |

Таблица 4. Результат экспертных оценок Эксперта 3

| | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 |
|-----|------|------|------|------|------|------|------|------|------|------|
| F1 | 1,00 | 5,00 | 1,00 | 5,00 | 5,00 | 1,00 | 3,00 | 1,00 | 5,00 | 3,00 |
| F2 | 0,20 | 1,00 | 0,14 | 0,20 | 0,33 | 0,14 | 0,20 | 0,14 | 0,33 | 0,33 |
| F3 | 1,00 | 7,00 | 1,00 | 5,00 | 9,00 | 1,00 | 7,00 | 1,00 | 7,00 | 7,00 |
| F4 | 0,20 | 5,00 | 0,20 | 1,00 | 5,00 | 0,33 | 0,33 | 0,33 | 0,20 | 0,20 |
| F5 | 0,20 | 3,00 | 0,11 | 0,20 | 1,00 | 0,11 | 0,14 | 0,11 | 0,33 | 0,33 |
| F6 | 1,00 | 7,00 | 1,00 | 3,00 | 9,00 | 1,00 | 5,00 | 1,00 | 5,00 | 5,00 |
| F7 | 0,33 | 5,00 | 0,14 | 3,00 | 7,00 | 0,20 | 1,00 | 0,33 | 0,33 | 0,20 |
| F8 | 1,00 | 7,00 | 1,00 | 3,00 | 9,00 | 1,00 | 3,00 | 1,00 | 7,00 | 7,00 |
| F9 | 0,20 | 3,00 | 0,14 | 5,00 | 3,00 | 0,20 | 3,00 | 0,14 | 1,00 | 1,00 |
| F10 | 0,33 | 3,00 | 0,14 | 5,00 | 3,00 | 0,20 | 5,00 | 0,14 | 1,00 | 1,00 |

Для каждой таблицы мы посчитали коэффициенты согласованности, чтобы убедиться в корректности выставленных оценок. Приемлемо, если коэффициент согласованности не превышает значение 0.1. Мы получили следующие результаты: 0.1, 0.07, 0.09 для каждой из таблиц соответственно. Данные значения не противоречат заданному условию.

Убедившись в том, что оценки экспертов согласованы, мы произвели расчеты для получения весов каждого признака для трех таблиц.

Существует 4 способа подсчета весов. Для первых двух способов необходимо найти сумму элементов каждой строки (для первого способа) и каждого столбца (для второго способа) и нормализовать эти значения. Для третьего способа необходимо произвести усреднение по нормализованным столбцам. Однако данные подходы дают наименее точные оценки по сравнению с четвертым подходом, который заключается в том, что необходимо перемножить элементы каждой строки и извлечь корень n-й степени, где n – количество элементов в строке, а затем нормализовать полученные числа [1]. В данной работе мы придерживаемся последнего подхода.

Мы посчитали значения весов каждого эксперта, усреднили данные показатели и получили следующие значения (Таблица 5).

Таблица 5. Веса суггестивности признаков

| Признак | Вес |
|---------|---------------|
| F1 | 0,1621334385 |
| F2 | 0,02987648162 |
| F3 | 0,1555681428 |
| F4 | 0,07370519761 |
| F5 | 0,02989766824 |
| F6 | 0,1678753404 |
| F7 | 0,02678581843 |
| F8 | 0,2058984718 |
| F9 | 0,07182808343 |
| F10 | 0,0764313572 |

Таким образом, исходя из оценок экспертов, наиболее суггестивным является признак F8, то есть повелительное наклонение глаголов, а наименее суггестивным является F7 (устойчивые выражения).

Этап 3. Лингвистический анализ.

На данном этапе необходимо провести лингвистический анализ рекламных текстов, выявить 10 признаков суггестии и произвести расчет метрики суггестивности.

Этап 4. Расчёты.

Далее необходимо произвести расчеты метрики суггестивности по формуле (1):

$$\sum_{i=1}^n w_i \times \delta_i \delta_i = \begin{cases} 1, & \text{если признак } i \text{ есть в тексте} \\ 0, & \text{если признак } i \text{ отсутствует} \end{cases} \quad (1)$$

Этап 5. Сортировка полученных результатов.

На данном этапе необходимо отсортировать полученные оценки суггестивности для каждого текста в порядке убывания.

Этап 6. Выбор наиболее суггестивного текста.

Получив отсортированные оценки для текстов, необходимо выбрать первый текст, как наиболее суггестивный.

Таким образом, данный метод позволяет выбрать наиболее суггестивный текст из предложенных вариантов.

Эксперимент

Рассмотрим пример того, как можно применить разработанную нами методологию на практике.

В качестве примера рекламного текста, возьмем описание одного из товаров с маркетплейса Wildberries [3].

Универсальный фиксатор для бровей PÚSY Brow Fix, добро пожаловать в удивительный мир косметики и ухода за вашими бровями! Гель пуси для бровей с эффектом ламинирования 5 мл - идеальный выбор для надежной и долговременной фиксации. Невидимый при нанесении, этот гель поможет вам создать безупречную укладку бровей. Благодаря уникальному составу, включающему специальные питательные вещества, фиксатор для бровей способен питать и укреплять волосы, придавая им здоровый и ухоженный вид. Вы никогда не подозревали, насколько важна правильная фиксация и уход за бровями, пока не попробовали этот продукт. Фиксируйте свои брови с уверенностью!

Проведя лингвистический анализ, мы выявили следующие суггестивные средства:

1. Имена собственные: название бренда PÚSY;
2. Слова с конкретным значением, например, брови, гель, волосы;
3. Заимствованная лексика: универсальный, фиксатор;
4. Лексические повторы, например, фиксатор, универсальный, бровей, гель;

5. Числительные: 5 мл;

6. Повелительное наклонение глаголов: *фиксируйте*.

Таким образом, мы выявили следующие признаки: F1, F2, F3, F4, F6 и F8. Суммируя веса для данных признаков, мы получаем, что степень суггестивности рекламного текста 79.5%.

Заключение

В результате данного исследования мы разработали методологию, которая позволяет определить, насколько текст суггестивен, на основе экспертных оценок суггестивных признаков получили их веса значимости и продемонстрировали работу данного подхода.

Список литературы

1. Saati T. Decision-making. Method of the analysis of hierarchies //M.: Radio and communication. – 1993. – Т. 278.
2. Saaty T. L., Decision making with the Analytic Hierarchy Process //International Journal of Services Sciences. – 2008. – Vol. 1. – No. 1. – P. 83-98.
3. Wildberries [Электронный ресурс]. URL: <https://www.wildberries.ru/> (дата обращения: 18.02.2024).
4. Бернадская Ю. С. Текст в рекламе: учеб. пособие для студентов вузов. – М.: Юнити, 2008. – 287 с.
5. Бойчук Е. И. Ритм художественного текста как средство суггестивного воздействия //Ярославский педагогический вестник. – 2014. – Т. 1. – №. 2. – С. 129-133.
6. Голайденко Л. Н., Суханова Д. Н. Письменный текст в телевизионной рекламе: речевые средства воздействия на зрителя //Современное развитие науки: вопросы теории и практики: сборник материалов II-ой международной научно-практической конференции. – Москва. – 2016. – С. 15.
7. Селеменова О. А. Языковые средства создания эффекта доверия в коммерческих рекламных текстах //Вопросы теории и практики журналистики. – 2020. – Т. 9. – №.3. – С. 530-542.
8. Сидоренко Е. В. Тренинг влияния и противостояния влиянию. – СПб. – 2002. – С. 30-31.
9. Цуканова Е. М. Языковые средства выражения суггестии в рекламном слогане // Ученые записки Орловского государственного университета. Серия: Гуманитарные и социальные науки. – 2016. – №. 3 (72). – С. 234-239.
10. Шелестюк Е. В., Речевое воздействие: онтология и методология исследования: монография – 2-е изд., испр. и доп. – М.: ФЛИНТА: Наука, 2014. – 344 с.

ИССЛЕДОВАНИЕ И ПРИМЕНЕНИЕ ТЕХНОЛОГИИ VOLUME RENDERING ДЛЯ ВИЗУАЛИЗАЦИИ РЕЗУЛЬТАТОВ МРТ

*Башинский Д.О.¹, Шин В.А.², Семичаснов И.В.³
Национальный исследовательский университет
«Высшая школа экономики»,*

¹департамент компьютерной инженерии,

²департамент электронной инженерии,

*³Центр управления проектными разработками
МИЭМ НИУ ВШЭ*

Аннотация

Технологии виртуальной реальности активно применяются в медицине для подготовки и проведения операций. Эти технологии помогают создавать модели пациентов, обеспечивая хирургов информацией о возможных рисках и проблемах при сложных вмешательствах. В работе рассматривается применение технологии Volume Rendering для визуализации результатов магнитно-резонансной томографии (МРТ) в виртуальной реальности с использованием игрового движка Unity3D.

Целью данного исследования является изучение и реализация способов визуализации данных, полученных посредством компьютерно-томографического сканирования.

Авторами были описаны и продемонстрированы методы визуализации данных, полученных в результате МРТ сканирования. Также было создано приложение, позволяющее исследовать результаты МРТ-исследований в виртуальной среде и взаимодействовать с ней.

Введение

В настоящее время использование виртуальной реальности в медицине является одним из наиболее перспективных направлений, способствующих совершенствованию методов подготовки и проведения медицинских операций. Технологии виртуальной реальности открывают широкие возможности для создания точных трехмерных моделей пациентов и визуализации сложных медицинских данных, что значительно улучшает качество медицинской практики.

В работе рассматриваются вопросы применения технологии Volume Rendering для визуализации результатов магнитно-резонансной томографии в виртуальной реальности с использованием игрового движка Unity3D [1]. Основной задачей исследования является анализ результатов виртуальной визуализации.

Практическая значимость данного исследования заключается в возможности повышения качества подготовки к хирургическим вмешательствам, а также в улучшении понимания хирургами возможных рисков и сложностей, связанных с операциями, благодаря более детальной и реалистичной визуализации данных.

Научная новизна исследования определяется разработкой специализированного приложения для исследования виртуальных визуализаций МРТ данных и их интерактивного анализа, что может способство-

вать последующему развитию методов визуализации медицинских данных и их применению в практической медицине.

В настоящее время существуют программные решения, схожие по функционалу с нашей разработкой. К ним относятся 3D SLICER, AdvantageSim MD, MRICrop. Данные приложения успешно интегрированы в рабочие процессы медицинских учреждений.

Однако, наше решение имеет два преимущества перед вышеупомянутыми аналогами. Во-первых, оно предоставляет пользователям возможность выбора между двумя видами отображения данных в виртуальной среде: двумерным и трехмерным. Во-вторых, оно обладает функцией отображения данных посредством шлема виртуальной реальности, что позволяет осуществлять виртуальную манипуляцию с данными сканирования с помощью физических устройств ввода. Данная особенность является ключевым преимуществом нашего приложения перед существующими аналогами.

Технология Volume Rendering

Технология объемного рендеринга (Volume Rendering) представляет собой комплекс методов визуализации дискретных трехмерных (объемных) данных. Эти данные могут быть получены из различных источников, включая компьютерную томографию, магнитно-резонансную томографию, позитронно-эмиссионную и ультразвуковую томографию.

Существует множество применений Volume Rendering. В данном исследовании акцент сделан на использовании этого метода для визуализации сканов компьютерной томографии (КТ), которая может быть выполнена для диагностики травм, таких как переломы конечностей. Путем проведения компьютерной томографии формируется трехмерный объемный набор данных, в котором каждая ячейка содержит информацию о плотности соответствующего участка изображения.

Наборы объемных данных

В отличие от набора данных поверхности, объемный набор данных представляет собой трехмерное скалярное поле, значения которого определены в непрерывной трехмерной области.

Относительно объемного рендеринга, объемный набор данных хранится в виде трехмерного массива скалярных значений, которые получаются путем выборки непрерывной трехмерной области в этих дискретных местах.

Пиксель – это отдельный элемент изображения дискретного двумерного скалярного поля. На изображении элементы объема, составляющие объем выборки, называются вокселями. Можно представить, что объем состоит из совокупности однородных кубов, где каждая отдельная точка выборки находится в центрах соответствующих кубов. Хотя это упрощает визуализацию окружающего среды каждого отдельного вокселя, более точным подходом будет идентифицировать каждый отдельный воксель с вы-

боркой, полученной в бесконечно малой точке в R^3 . Обе модели представлены на рис.1.

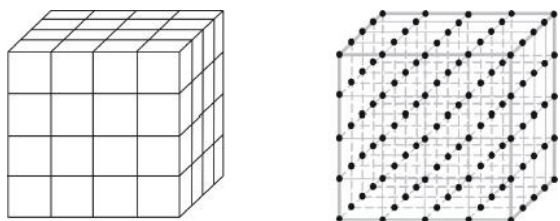


Рис.1. Модели набора объемных данных:

(1) Набор кубов одинакового размера, каждый из которых сосредоточен вокруг отдельной точки выборки;

(2) Набор скалярных значений, каждое из которых выбрано в бесконечно малых точках выборки

В работе используются трехмерные воксельные данные (рис.3), полученные в результате медицинского компьютерного томографического сканирования. Значение, хранящееся в каждой точке данных, отражает плотность материала исследуемой области. Эти данные извлекаются из файла и сохраняются в трехмерной текстуре. Ввиду удобства помещения данных в прямоугольный параллелепипед, создается соответствующая пространственная структура (отображаются лишь задние грани), а для передачи плотности каждого видимого вокселя внутри параллелепипеда применяется техника Raymarching (рис.2).

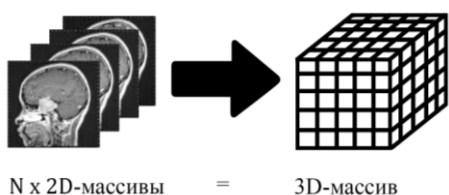


Рис.2. Конвертация данных КТ скана

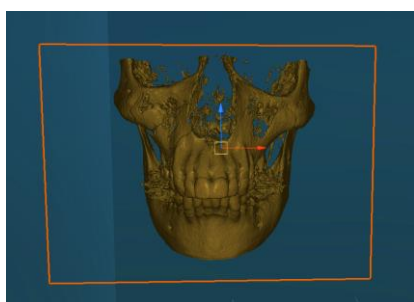


Рис.3. Визуализация данных с помощью технологии Volume Rendering

Методы Volume Rendering

На данный момент изучено множество различных способов рендеринга трехмерных скалярных полей, то есть объемных данных. В общем, эти методологии могут быть систематизированы с применением одного из следующих подходов к классификации.

1. Direct Volume Rendering

Метод прямого объемного рендеринга (Direct Volume Rendering) формирует изображения трехмерного скалярного объема без акцента на извлечение поверхности, соответствующей определенным интересующим объектам или определенной изовеличиной. Для непосредственного отображения данных, хранящихся в объемном наборе, необходима оптическая модель, описывающая взаимодействие скалярных значений объема со светом, например, излучение, поглощение, отражение или преломление [2].

Реализация данного метода [3]: на каждом этапе происходит слияние текущего цвета вокселя с накопленным значением цвета из предыдущих этапов. Результат визуализации приведен на рис.4.

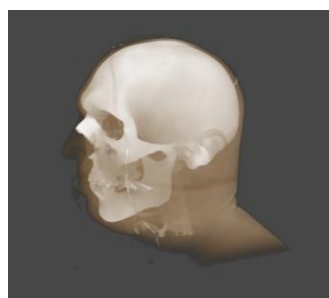


Рис.4. Результат визуализации с помощью Direct Volume Rendering

2. Maximum Intensity Projection

В методе Maximum Intensity Projection (MIP) цвет каждого пикселя в окончательной визуализации определяется на основе максимального значения обнаруженного при пропускании каждого соответствующего луча [4].

Применение данного метода рендеринга становится значимым в области обработки трехмерных медицинских наборов данных [5], полученных с использованием магнитно-резонансного томографа (МРТ). MIP акцентирует внимание на костях и заметных различиях в плотности, хотя в то же время не обеспечивает точного извлечения поверхностей, соответствующих определенным органическим тканям. Результат визуализации приведен на рис.5.

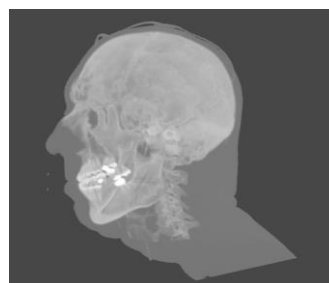


Рис.5. Результат визуализации с помощью Maximum Intensity Projection

3. *Isosurface Rendering*

В объемном рендеринге изоповерхность (isosurface) характеризует контурную поверхность, извлеченную из набора объемных данных, которая соответствует определенному объекту. Она определяется по значению изовеличины (isovalue), связанной с этим объектом. Термин изоповерхности также используется для обобщения граничных поверхностей областей объемных данных с однородными атрибутами. Например, изоповерхность может описывать область с плотностью выше или ниже заданного порога.

Изоповерхность состоит из множества полигонов (многоугольников), определяющих ее геометрию и извлекаемых во время предварительной обработки перед отображением. На сегодняшний день существует множество алгоритмов для извлечения изоповерхностей из объемных данных, например, алгоритм Marching Cubes [6], который используется для генерации геометрического представления объекта.

Изоповерхности могут быть визуализированы с помощью метода изоповерхностного рендеринга (Isosurface Rendering) [7]. В этом случае они называются неполигональными изоповерхностями. Для отображения неполигональных изоповерхностей используется метод бросания лучей (ray-casting) с использованием специальных передаточных функций [8]. Результат визуализации приведен на рис.6.



Рис.6. Результат визуализации с помощью *Isosurface Rendering*

Способы улучшения визуализации сформированных объектов

Применение методов объемного рендеринга дало возможность создать визуализацию трехмерных данных. Путем использования техник расчета освещения и передаточных функций достигается улучшение визуальных результатов, что способствует повышению наглядности и понятности восприятия визуализации.

1. Освещение

Для вычисления освещения применяется метод диффузного отражения света [9]. На рис.7 представлен результат визуализации с помощью *Isosurface Rendering* и применением метода диффузного отражения света.

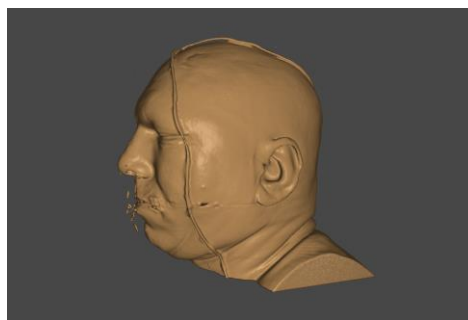


Рис.7. Результат визуализации с помощью *Isosurface Rendering*

2. Передаточная функция

Использование передаточных функций позволяет определять цвет и прозрачность вокселя на основе его плотности [10]. На рис.8 продемонстрирован результат визуализации с использованием метода *Direct Volume Rendering* и применением двумерной передаточной функции. На данном изображении сделан акцент на костных слоях.

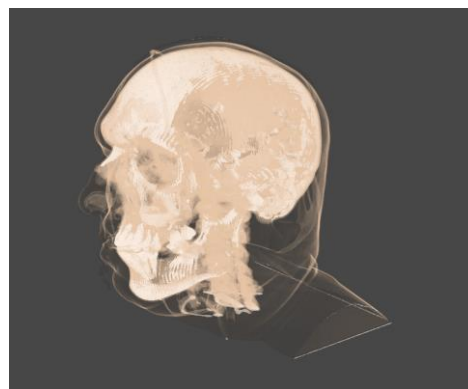


Рис.8. Результат визуализации с помощью *Direct Volume Rendering* и двумерной передаточной функции

Взаимодействие пользователя с полученными объектами визуализации

Для обеспечения удобного и эффективного взаимодействия с данными, полученными в результате визуализации, было разработано приложение на движке *Unity3D* в сфере виртуальной реальности. На рис.9 представлены скриншоты разработанного VR-приложения на движке *Unity3D*.

Приложение позволяет пользователю манипулировать и взаимодействовать с визуализированными объектами при помощи контроллеров.

Возможности пользователя:

- Изменение точки обзора;
- Перемещение в виртуальном пространстве;
- Выбор объектов визуализации из имеющегося набора;
- Перемещение и вращение плоскости сечения визуализированного объекта для более детального изучения конкретных областей объекта.

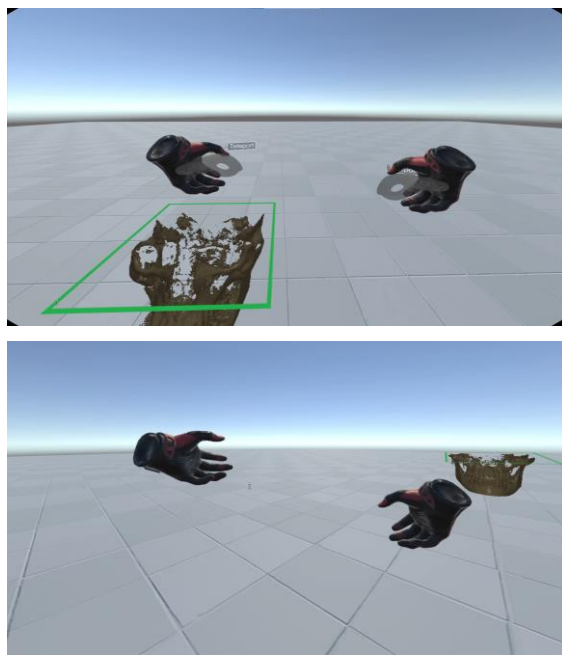


Рис.9. VR-приложение на движке Unity3D

Заключение

В данном исследовании был рассмотрен набор методов визуализации объемных данных, полученных с использованием магнитно-резонансной томографии. Для каждого из методов были определены процедуры визуализации трехмерных данных и представлены результаты визуализации. Кроме того, было разработано приложение виртуальной реальности, способное отображать данные с МРТ-сканирования и обеспечивать пользовательское взаимодействие с ними.

Список литературы

1. Cosmina I. A Glance into Virtual Reality Development Using Unity // Informatica Economica, 2018.
2. N. Max, Optical models for direct volume rendering, IEEE Transactions on Visualization and Computer Graphics, 1995.
3. M. Meisner, J. Huang, D. Bartz, K. Mueller, and R. Crawfis, A Practical Evaluation of Popular Volume Rendering Algorithms, Volume Visualization, 2000.
4. Luccichenti, Giacomo, From Maximum Intensity Projection to Volume Rendering, Seminars in Ultrasound Ct and Mri, 2001.
5. Perandini S, Faccioli N, Zaccarella A et-al. The diagnostic contribution of CT volumetric rendering techniques in routine practice. Indian J Radiol Imaging. 2010.
6. W. E. Lorensen and H. E. Cline, Marching Cubes: A high resolution 3D surface construction algorithm, SIGGRAPH, 1987.
7. Fei Yang, Yong Cao, Jie Tian Efficient and Effective Volume Visualization with Enhanced Isosurface Rendering, 2012.
8. M. Levoy, Display of surfaces from volume data, IEEE Computer Graphics and Applications, 1988.

9. L. B. Wolff, S. Nayar, Michael Oren Improved Diffuse Reflection Models for Computer Vision // International Journal of Computer Vision, 1998.

10. Kalarat, Kosin & Koomhin, Phanit, Real-time Volume Rendering Interaction in Virtual Reality, International Journal of Technology, 2019.

11. Zhou, Jianlong & Tönnies, Klaus. 1 State of The Art for Volume Rendering.

12. J. Kniss, G. Kindlmann, and C. Hansen, Multi-Dimensional Transfer Functions for Volume Rendering, Scientific Computing and Imaging Institute, School of Computing, University of Utah, Johnson/Hansen: The Visualization Handbook, 2004.

13. P. Lücke, Volume Rendering Techniques for Medical Imaging, in collaboration with Siemens Corporate Research Inc., Princeton, USA, 2005.

14. P. Herghelegiu, M. Gavrilescu and Vasile Manta, Visualization of Segmented Structures in 3D Multimodal Medical Data Sets, Advances in Electrical and Computer Engineering, 2011.

15. R. Drebin, L. Carpenter & P. Hanrahan, Volume Rendering. SIGGRAPH Computer Graphics and Interactive Techniques, 1988.

16. Fishman, Elliot & Magid, D & Ney, D & Chaney, E & Pizer, S & Rosenman, Julian & Levin, D & Vannier, Michael & Kuhlman, J & Robertson, Douglas, Three-dimensional Imaging, Radiology, 1991.

КЛЮЧЕВЫЕ АСПЕКТЫ ОЦЕНКИ КАЧЕСТВА СИСТЕМ ГЕНЕРАЦИИ С ДОПОЛНЕННОЙ ВЫБОРКОЙ

Жариков Е.И.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент фундаментальной и
прикладной лингвистики
ФГН НИУ ВШЭ*

Аннотация

В данной работе исследуется вопрос автоматической проверки качества выдачи систем генерации с дополненной выборкой. Рассматриваются различные метрики, применяемые в этих системах, а также их различия и особенности. Далее рассматриваются открытые программные решения для оценки качества. В результате исследования выявляются ключевые аспекты и особенности оценки качества систем генерации с дополненной выборкой, что может способствовать улучшению эффективности и точности этих систем.

Введение

С увеличением популярности больших языковых моделей (LLM) многие исследователи и организации стремятся использовать их возможности в различных приложениях. Однако, задача расширения области знаний уже обученной модели является непростой. Отвечая на вопросы, на основе которых обучение не происходило, она может давать неверные ответы или

отвечать «галлюцинациями» [1]. В таких случаях метод Retrieval Augmented Generation для достижения лучших результатов является актуальным.

Retrieval Augmented Generation (RAG) [2] – это метод генерации ответа с учетом дополнительной информации. Таким образом, когда пользователь отправляет запрос в большую языковую модель она обеспечивает его информацией не на основе данных, на которых она обучалась, а на релевантных данных, полученных из заранее сформированной базы данных. Это повышает актуальность ответа и таким образом снижается шанс «галлюцинаций».

Однако, несмотря на потенциальные преимущества RAG, оценка качества таких систем остается актуальной проблемой. Изначально генерацию с дополнительной выборкой оценивали с точки зрения самой задачи языкового моделирования, но, к сожалению, такие методы не всегда позволяют полноценно оценить работу генерации [3]. Именно поэтому данная работа посвящена ключевым аспектам оценки качества систем генерации с дополнительной выборкой.

Открытые программные решения для подсчета метрик

Для оценки эффективности RAG существуют различные открытые программные решения:

1. Ragas [4];
2. Promptfoo [5];
3. RaLLe [6].

В данной статье рассматривается фреймворк RAGAS, ниже упомянуты несколько причин, почему выбран именно этот инструмент для оценки RAG.

1. **Обширный функционал.** RAGAS предоставляет широкий набор функций и инструментов для оценки и анализа работы RAG моделей. Он позволяет проводить эксперименты с различными параметрами, настраивать модели и проводить сравнительный анализ результатов.

2. **Простота использования.** RAGAS разработан с учетом удобства использования и предоставляет простой и интуитивно понятный интерфейс для проведения оценки RAG моделей. Это позволяет исследователям и разработчикам быстро и эффективно проводить эксперименты и анализировать результаты.

3. **Активное сообщество.** RAGAS имеет активное сообщество пользователей и разработчиков, которые активно вносят свой вклад в развитие инструмента. Это обеспечивает доступ к обсуждениям, руководствам и поддержке со стороны сообщества, что помогает пользователю получить максимальную отдачу от использования RAGAS.

Однако, у данного фреймворка есть и ограничения. Изначально, чтобы сразу использовать этот фреймворк требуется API ключ OpenAI так как некоторые метрики измеряются с помощью других больших языковых моделей, например GPT-4. Важно учесть, что иметь ключ не обязательно, так как данный фреймворк дружелюбен к другим открытым моделям. Главное – это соединить требуемую модель с программным интерфейсом RAGAS.

Метрики RAGAS

Существует большое количество метрик, с помощью которых можем оценить работу различных частей нашей работы. Некоторые метрики оценивают качество контекста, подобранного под вопрос, а другие метрики, например, сравнивают результат модели и истинный ответ. Перед тем как перейти к метрикам, стоит уточнить, что для снятия метрик есть 4 элемента, которые участвуют в оценке RAG:

- вопрос – запрос, который был задан пользователем;
- контекст – подобранный текст под вопрос пользователя из базы данных;
- истина – истинный ответ, который в идеальном случае, должна дать модель;
- ответ модели – результат, полученный от модели после того, как она получила вопрос и контекст.

После ознакомления с основными элементами рассмотрим метрики:

1. **Правдивость (Faithfulness)** С помощью данной метрики измеряется фактическая согласованность сгенерированного ответа модели с предоставленным контекстом. Значение по данной метрике находится в диапазоне [0,1], где чем ближе число к 1, тем правдивее является ответ. Сгенерированный ответ считается правдивым, если все утверждения, сделанные в ответе, могут быть выведены из предоставленного контекста.

2. **Релевантность ответа (Answer Relevancy)** Метрика оценки релевантности ответа фокусируется на оценке того, насколько соответствует сгенерированный ответ заданному вопросу. Более низкий показатель присваивается ответам, которые являются неполными или содержат избыточную информацию. Эта метрика рассчитывается на основе вопроса и ответа. Значение по данной метрике находится в диапазоне [0,1], где более высокие значения указывают на более высокую релевантность.

3. **Точность контекста (Context Precision)** Метрика точности контекста позволяет оценить, насколько все релевантные элементы из исходного контекста расположены на более высоких позициях. Эта метрика рассчитывается на основе вопроса и контекста. Значение по данной метрике находится в диапазоне [0,1], где значения, которые находятся ближе к 1, указывают на более высокую точность.

4. **Релевантность контекста (Context Relevancy)** Эта метрика позволяет оценить соответствие извлеченного контекста, рассчитанное на основе как вопроса, так и контекста. Значение по данной метрике находится в диапазоне [0,1], где более высокие значения указывают на более высокую релевантность. Идеально, извлеченный контекст должен содержать только существенную информацию для ответа на заданный вопрос.

5. **Полнота контекста (Context Recall)** Метрика полноты контекста измеряет степень соответствия извлеченного контекста аннотированному ответу, который рассматривается как истинное значение. Она рассчитывается на основе истинного значения и извлеченного контекста. Значение по данной метрике находится в диапазоне [0,1], где более высокие значения указывают на более высокую производи-

тельность. Для оценки полноты контекста из истинного значения ответа анализируется каждое предложение, чтобы определить, можно ли его отнести к извлеченному контексту или нет. В идеальном случае все предложения в истинном значении ответа должны быть отнесены к извлеченному контексту.

6. Семантическая схожесть ответа (Answer Semantic Similarity) Концепция семантической схожести ответа связана с оценкой семантического сходства между сгенерированным ответом и истинным значением. Эта оценка основывается на истинном значении и ответе. Значение по данной метрике находится в диапазоне [0,1], где более высокий показатель указывает на более высокую соответствие между сгенерированным ответом и истинным значением. Измерение семантической схожести между ответами может предоставить ценную информацию о качестве сгенерированного ответа. Для этой оценки используется модель перекрестного кодировщика для расчета показателя семантической схожести.

7. Правильность ответа (Answer Correctness) Оценка правильности ответа включает определение точности сгенерированного ответа по сравнению с истинным значением. Эта оценка основывается на истинном значении и ответе модели. Значение по данной метрике находится в диапазоне [0,1], где чем число ближе к 1, тем ответ является более схожим с истинным значением. Правильность ответа включает два важных аспекта: семантическое сходство между сгенерированным ответом и истинным значением, а также фактическое сходство.

Лингвистические метрики

В нашем исследовании были отобраны следующие метрики для оценки работы RAG с лингвистической точки зрения:

1. неверный выбор слова;
2. неверный порядок слов;
3. неверная расстановка знаков препинания;
4. неверное построения ряда однородных членов;
5. неудачное введение противопоставления;
6. неудачная грамматическая конструкция;
7. неверное употребление местоимений;
8. пропуск смыслового звена;
9. дефект мысли;
10. употребление неизвестных слов без должного объяснения;
11. употребление слова одновременно в двух значениях.

Данные метрики основаны на учебном пособии для журналистов для анализа речи [8]. В нём рассматриваются вышеупомянутые смысловые ошибки. Так как автоматические метрики фреймворка RAGAS находятся в диапазоне от 0 до 1, то лингвистические метрики тоже будут ранжироваться от 0 (имеется ошибка) до 1 (данной ошибки в тексте нет).

Открытые программные решения для разворачивания модели

Существуют различные open-source модели, которые можно использовать для измерения метрик в задачах RAG. Обычно, для использования этих моделей требуются сторонние библиотеки, чтобы

развернуть их на локальном сервере. RAGAS совместим с многими программными решениями.

Библиотека vllm [8] предоставляет инструменты и функционал для работы с языковыми моделями, но она требует запуск на операционной системе Linux и использования графического процессора (GPU). Однако, для пользователей, предпочитающих операционную систему Windows, существует также библиотека Anyscale [9], которая тоже предоставляет широкий функционал по работе с большими языковыми моделями.

Рекомендуется использовать модели с числом параметров, составляющих порядка 7 миллиардов, для получения качественных ответов. Для использования этих моделей достаточно просто развернуть локальный сервер с моделью и интегрировать ее в систему с помощью соответствующих интерфейсов. Это обеспечивает возможность проведения оценки и измерения метрик с использованием мощных открытых моделей в задачах RAG.

Оценка с помощью автоматических метрик

Для сбора данных в рамках данного исследования были использованы документы, собранные с официального сайта Высшей школы экономики в Нижнем Новгороде, относящиеся к теме «поступление». Данные документы представляют собой информацию о процедурах и правилах приёма абитуриентов в данное образовательное учреждение. Всего было получено 9 таких документов. База данных, с помощью которой будут представляться релевантные данные, заполняется фрагментами из документов по 150-200 символов. Также для каждого документа было разработано по 5 вопросов, с целью выявить определенные аспекты и информацию, необходимую для дальнейшего анализа. Каждый вопрос был составлен с учетом содержания соответствующего документа и его важности для исследования.

Полученный материал оценивался с помощью автоматических метрик. В текущем исследовании предполагается использовать вышеупомянутый фреймворк RAGAS и метрики, которые он предоставляет для оценки (Таблица 1).

Таблица 1.

| Модель | Правильность ответа | Правдивость | Релевантность ответа |
|----------------|---------------------|-------------|----------------------|
| RuGPT 3-small | 0.53 | 0.47 | 0.48 |
| RuGPT 3-medium | 0.61 | 0.48 | 0.52 |
| RuGPT 3-large | 0.63 | 0.62 | 0.59 |

Из результатов оценки моделей с помощью фреймворка RAGAS можно сделать следующие выводы:

1. Модель RuGPT3-large показала наилучшие результаты по всем рассмотренным метрикам: высокая правильность ответов (0.63), правдивость (0.62) и релевантность ответов (0.59). Это говорит о том, что данная модель способна генерировать более точные, правдивые и релевантные ответы.

2. Модели RuGPT3-medium и RuGPT3-small также показали хорошие результаты, однако их показатели немного ниже, чем у модели RuGPT3large.

Таким образом, анализ результатов оценки моделей с помощью фреймворка RAGAS позволяет сделать вывод о том, что модель RuGPT3-large демонстрирует лучшие показатели среди рассмотренных моделей, обладая высокой правильностью, правдивостью и релевантностью ответов среди моделей RuGPT3-large, RuGPT3-medium и RuGPT3-small.

Заключение

В работе рассматриваются существующие системы оценивания RAG (Retrieval Augmented Generation), которые используются для оценки качества систем генерации с дополненной выборкой. Были рассмотрены различные метрики, как с математической, так и с лингвистической точек зрения, которые применяются в этих системах, и обсуждены их различия и особенности. Также были изучены открытые программные решения для оценки качества и модели, которые могут быть использованы для оценки. В результате исследования были выявлены ключевые аспекты и особенности оценки качества систем генерации с дополненной выборкой. Это знание может способствовать улучшению эффективности и точности этих систем. Также был произведен анализ с помощью автоматических метрик и были выявлены модели с более лучшими показателями. Дальнейшие исследования и разработки в этой области могут привести к улучшению систем оценивания RAG и, как следствие, к более качественной генерации с дополненной выборкой.

Список литературы

1. Marcus G. The next decade in AI: four steps towards robust artificial intelligence //arXiv preprint arXiv:2002.06177. – 2020.
2. Lewis P. et al. Retrieval-augmented generation for knowledge-intensive nlp tasks //Advances in Neural Information Processing Systems. – 2020. – Т. 33. – С. 9459-9474.
3. Wang J. et al. Is chatgpt a good nlg evaluator? a preliminary study //arXiv preprint arXiv:2303.04048. – 2023.
4. Ragas [Электронный ресурс]. URL: <https://docs.ragas.io/en/stable/index.html> (дата обращения: 15.11.2023)
5. Promptfoo [Электронный ресурс]. URL: <https://github.com/promptfoo/promptfoo> (дата обращения 13.02.2023)
6. RaLLe [Электронный ресурс]. URL: <https://github.com/yhoshi3/RaLLe?tab=readme-ov-file> (дата обращения 13.02.2023)
7. Асташова О. И., Колясникова О. С. Критика речи: смысловые ошибки: практикум. – 2016.

8. Vllm [Электронный ресурс]. URL: <https://github.com/vllm-project/vllm> (дата обращения 13.02.2023)

9. Anyscale [Электронный ресурс]. URL: <https://www.anyscale.com> (дата обращения 13.02.2023)

ПРЕДСКАЗАНИЕ КОММЕРЧЕСКОЙ ЭФФЕКТИВНОСТИ ВИДЕОИГР НА ОСНОВЕ АНАЛИЗА КОММЕНТАРИЕВ МЕТОДАМИ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

*Захаров Ф.А.¹, Молоканов Р.А.¹,
Литвиненко Е.И.¹, Малышкина К.А.²*
*Национальный исследовательский университет
«Высшая школа экономики»,
¹департамент прикладной математики,
²департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе изучается возможность эффективного предсказания коммерческих показателей видеоигр на рынке на основе мультимодальных текстовых данных, описывающих их: комментариев игроков и отзывов критиков.

Авторами с использованием механизмов веб-скраппинга был сформирован датасет, содержащий описанную выше информацию для 13117 различных видеоигр. Для прогнозирования рыночных показателей были использованы методы обработки естественного языка на основе искусственных нейросетей. В том числе был предложен новый подход, основанный на использовании факта разнородности данных, описывающих игру. Он был сопоставлен с существующими методами на основе ряда метрик оценки качества регрессии.

Введение

Задача построения качественной коммерческой аналитики в сфере геймдизайна (разработки игр) является на данный момент предельно актуальной. Ведь сегодня не существует зарекомендовавших себя математических моделей, позволяющих проанализировать шансы игры на успех на ранних этапах разработки, когда в нее еще не вложены колоссальные денежные средства и есть возможность скорректировать процесс ее создания. На данный момент анализ потенциальных коммерческих успехов игры в индустрии проводится сугубо на экспертном уровне, без использования каких-либо математических моделей и алгоритмов. Авторами исследования была выдвинута гипотеза о том, что рыночные показатели популярности игры могут быть спрогнозированы на основе текстовых данных о ней, а именно – отзывов обычных игроков и рецензий профессиональных критиков (которые могут быть собраны как раз на ранних этапах при тестировании демоверсии на ограниченной аудитории).

Основной теоретический и методологический упор в работе был сделан на изучении и использовании в разработке собственных алгоритмов современных методов обработки естественного языка. Одной из наиболее эффективных на сегодняшний день моделей для построения численного представления текстового корпуса является BERT [1] (Bidirectional Encoder Representations from Transformer), построенная на основе архитектуры типа Transformer [2]. Значимость BERT для аналитики естественного языка подтверждается множеством источников [3-5]. Помимо этого, BERT успешно применяется в различных приложениях и специфичных задачах [6-7]. Также предварительно были исследованы методы комбинирования информации о нескольких объектах, вследствие чего обнаружилось, что некоторые методы мультимодальности [8] не подходят из-за переменного количества текстов. Для решения этой проблемы авторы рассматривают подходы, аналогичные обработке рекуррентных данных, такие как LSTM (Long Short-Term Memory) [9], которые позволяют получить численное представление набора объектов путем последовательной их обработки.

В результате исследования было разработано пять нейросетевых архитектур (две из которых основаны на использовании мультимодальности данных), которые были протестированы на сформированном авторами датасете. Также стоит отметить, что помимо текстовых корпусов были использованы и доступные численные метаданные (дата и сопутствующие оценки по шкале от 0 до 10 для отзывов и рецензий).

Описание данных и формирование датасета

Поскольку не было найдено научных исследований, посвященных мультитекстовой модальности, авторы решили собрать информацию самостоятельно. Был сформирован датасет в соответствии со следующими требованиями: наличие различного количества текстов для каждого объекта; есть возможность выделения кластеров текстов с похожими свойствами (называемых "типами текстов"); есть наличие метаинформации для каждого текста. В результате авторами представлен новый датасет «GAMEFUSE», содержащий информацию о 13117 играх. Каждая игра описывается двумя видами текстов: комментариями игроков и рецензиями критиков. Комментарии представляют собой эмоциональные, неструктурированные тексты, выражающие мнения игроков, в то время как рецензии - это полноценные субъективные тексты, описывающие как положительные, так и отрицательные аспекты игры. Каждому комментарию присвоена оценка от 0 до 10, выставленная пользователем, и указана дата публикации. Для рецензий доступна только дата публикации. Итоговый объем собранных данных составляет 1,5 миллиона комментариев и 90 тысяч рецензий.

Покажем, что тексты разных типов (комментарии и рецензии) являются различными не только с интуитивной, но и с математической точки зрения. Доказанная авторами разнородность данных может быть

эффективно использована при построении репрезентативного векторного представления игры.

Для этого были рассчитаны определенные статистические показатели для всех текстов этих типов, описывающие семантические, стилистические, синтаксические, лексикографические особенности. Эти показатели являются информативными и часто используются для определения авторства текстов [10-12]. Для каждого текста было подсчитано 10 показателей: средняя длина слова, средняя длина предложения, среднее количество вопросительных предложений, относительное количество слов длиннее 4 букв, относительное количество знаков препинания и другие. Затем 10-мерные векторы были спроецированы в двухмерное пространство с помощью алгоритма t-SNE [13]. Изображение, демонстрирующее результаты для десятой части объектов датасета, приведено ниже.

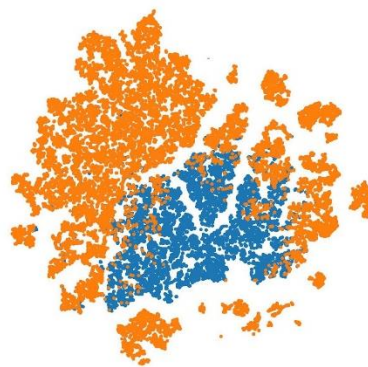


Рис. 1. Проекция наборов статистических характеристик текстов на двумерную плоскость: синий цвет – данные типа «рецензия», оранжевый цвет – данные типа «комментарий»

Основным источником данных стал портал «Metacritic» - крупнейший агрегатор пользовательских отзывов на игры в мире. Коммерческие показатели были взяты с онлайн-сервиса цифрового распространения компьютерных игр «Steam». В качестве них выступили верхняя и нижняя грани количества покупок на портале, а также количество подписок на новости об игре.

Предложенные модели и методы анализа собранных данных

Авторами было реализовано пять нейросетевых моделей (три из них используют классические подходы агрегации текстовой информации, последние две представляют инновационное решение, используя мультимодальность входных данных). Программная реализация производилась на языке Python с использованием фреймворка PyTorch. Ниже приведено более детальное описание моделей. Они упорядочены по возрастанию сложности и предполагаемой описательной способности (до тестирования в рамках эксперимента):

BERT + avg:

Для получения векторного представления каждого из текстов используется BERT. Далее каждое век-

торное представление (эмбединг) обрабатывается двумя полносвязными слоями нейронной сети с функцией активации ReLU для получения численного ответа по трем целевым рыночным показателям. Для получения итогового результата игры показатели, предсказанные для каждого из текстов, усредняются.

stacking + BERT:

Отличие второй модели от первой в том, что она не обрабатывает каждый текст отдельно, а конкатенирует все тексты и пропускает через вышеописанный пайплайн единый текстовый корпус, получая один ответ.

BERT + LSTM:

Данная модель после получения эмбедингов каждого из текстов обрабатывает их последовательно с помощью архитектуры LSTM, получая векторное описание игры. Далее этот вектор обрабатывается полносвязными слоями для получения численных ответов.

2*(BERT + LSTM):

Четвертая модель представляет из себя новое решение, основанное на обоснованной мультимодальности входных данных. Тексты каждого из двух типов обрабатываются обособленными архитектурами BERT и LSTM для получения эмбединга комментариев и эмбединга рецензий, которые далее конкатенируются и обрабатываются полносвязными слоями для получения итогового ответа.

2*(BERT + metainfo + LSTM):

Данная модель является улучшенной версией вышеописанной, использующей также метаданные о текстах. Метаданные конкатенируются к векторным представлениям текстов для дальнейшей обработки. Архитектура модели представлена на схеме ниже.

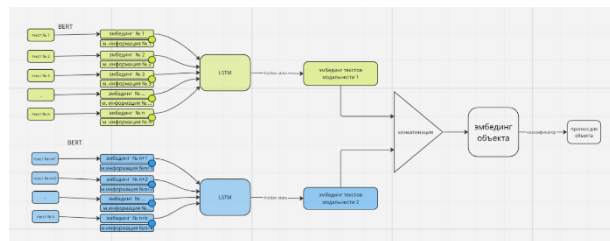


Рис.2. Модель «2*(BERT + metainfo + LSTM)», основанная на использовании мультимодальности текстов

Эксперимент и полученные результаты

Напомним, что в эксперименте в качестве целевых признаков для предсказания использовались коммерческие данные об игре, включая нижнюю и верхнюю оценки количества покупок в сервисе «Steam» (точное количество неизвестно из-за особенностей статистики сервиса) и количество подписчиков игры в «Steam». Оба показателя являются положительными целыми числами.

Стоит отметить, что данные о количестве покупок и подписок распределены неоднородно, при этом показатели очень популярных игр существенно превышают показатели большинства игр, зачастую на 3-

4 порядка. Вследствие этого авторы решили работать с этими признаками в логарифмической шкале. Данная техника была применена на основании опыта, упомянутого в работе [14].

В качестве последнего слоя нейросетевых моделей в рамках данного эксперимента авторами был использован полносвязный линейный слой с тремя выходами и линейной функцией активации. Функция средней абсолютной ошибки с суммированием по трем компонентам целевого вектора применялась в качестве функции потерь для обучения архитектур. Для оптимизации функции, использовался алгоритм градиентной оптимизации «Adam». Качество моделей оценивалось на основании следующих метрик:

$$\bullet MAE = \frac{\sum_{i=1}^n |y - \hat{y}_i|}{n}$$

$$\bullet MAPE = \sum_{i=1}^n \left| \frac{y - \hat{y}_i}{\hat{y}_i} \right| \times 100\%$$

Результаты эксперимента представлены в таблице ниже.

Таблица 1. Метрики качества моделей на обучающей и тестовой подвыборках датасета

| № | model | train MAPE | train add MAE | test MAPE | test MAE |
|---|-------------------------|---------------|---------------|---------------|--------------|
| 1 | BERT + avg | 0.3257 | 2.249 | 0.5179 | 2.135 |
| 2 | stacking+BERT | 0.3433 | 2.731 | 0.4702 | 2.805 |
| 3 | BERT + LSTM | 0.3634 | 2.19 | 0.3746 | 2.462 |
| 4 | 2*BERT + LSTM | 0.2995 | 2.056 | 0.3576 | 2.088 |
| 5 | 2*BERT+metainfo. + LSTM | 0.2629 | 1.992 | 0.3389 | 2.076 |

В серии экспериментов было установлено, что модель №5 превзошла остальные, что подтверждает изложенные выше гипотезы авторов об эффективности использования разнородности данных. Успех модели №5 по сравнению с моделью №4 указывает на важность учета данных о дате публикации текста и численной оценке комментатора при прогнозировании популярности игры.

Заключение

В этом исследовании был предложен ряд алгоритмов машинного обучения, решающих задачу предсказания коммерческого успеха видеоигр на основе отзывов игроков и рецензий критиков. Наряду с классическими подходами к агрегации текстовой информации были разработаны новые модели, использующие факт мультимодальности входных данных. Прделанная работа является серьезным шагом в построении автоматизированной аналитики коммерческого успеха видеоигр, адаптированной под специфику игровой индустрии. Ранее задача такого типа решалась исключительно вручную на ос-

нове субъективного опыта специалиста, что не могло гарантировать высокого качества результата и претендовать на масштабируемость. Эффективность предложенных методов и справедливость выдвинутых гипотез были подтверждены серией экспериментов на собранном авторами датасете.

Список литературы

1. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. //arXiv preprint arXiv:1810.04805 – 2019.
2. Vaswani, A., Shazeer N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., Polosukhin, I.: Attention Is All You Need //Advances in neural information processing systems – 2017. – Т. 30
3. Rogers, A., Kovaleva, O., & Rumshisky, A.: A primer in BERTology: What we know about how BERT works //Transactions of the Association for Computational Linguistics – 2021. – Т. 8 – С. 842-866
4. Singh, S., & Mahmood, A.: The NLP cookbook: modern recipes for transformer based deep learning architectures //IEEE Access – 2021. – Т. 9 – С. 68675-68702
5. Koroteev, M.: BERT: A Review of Applications in Natural Language Processing and Understanding. //arXiv preprint arXiv:2103.11943 – 2021.
6. Ghasemi, N., Aliannejadi, M., Hiemstra, D.: BERT for Target Apps Selection: Analyzing the Diversity and Performance of BERT in Unified Mobile Search. //arXiv preprint arXiv:2109.06306 – 2021.
7. Aldahdooh, J., Vähä-Koskela, M., Tang, J., Tanoli, Z.: Using BERT to identify drug-target interactions from whole PubMed – 2021.
8. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.: Multimodal Deep Learning //Proceedings of the 28th International Conference on Machine Learning ICML 2011 – 2011. – С. 689-696
9. Hochreiter, S., Schmidhuber, J.: Long Short-term Memory //Neural computation – 1997. – Т. 9 – С. 1735-80
10. Pervaz, I., Ameer, I., Sittar, A., Nawab, R.M.: Identification of Author Personality Traits using Stylistic Features //Conference and Labs of the Evaluation Forum – 2015.
11. Pennebaker, J., Mehl, M., Niederhoffer, K.: Psychological Aspects of Natural Language Use: Our Words, Our Selves //Annual review of psychology – 2003. – С. 547-77
12. Белов А.В., Егорова Э.А.: Применение методов ИИ в системе поддержки проведения научно-технических экспертиз// Интернет изнутри, 2023, № 18, С. 25-29
13. van der Maaten, L., Hinton, J.: Visualizing Data using t-SNE //Journal of Machine Learning Research – 2008. – Т. 9 – С. 2579-2605
14. Bellégo C., Benatia D., Pape L.: Dealing with logs and zeros in regression models //arXiv preprint arXiv:2203.11820 – 2022.

ДЕДУКТИВНО-НЕЙРОННЫЙ ПОДХОД НА ПРИМЕРЕ ФОРМАЛИЗАЦИИ РЕШЕНИЯ ЗАДАЧ ПЛАНИМЕТРИИ

Просвирнин Р.О.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ

Аннотация

В работе представлен новый метод решения математических задач с применением искусственного интеллекта – дедуктивно-нейронный подход (DNA). Рассмотрена структура DNA на примере задачи доказательства теорем планиметрии и предложен новый сервис DNASolver, основанный на системах AlphaGeometry [1] и "Нейрогеометрия" [2].

Введение

Автоматизация процесса доказательства теорем геометрии является давней и сложной задачей. Подходы последних лет были все ещё основаны на предикатных языках и эвристиках [3]. В силу ограниченности таких методов, нельзя было считать этот вопрос разрешённым. Отметим, что опыты над генеративными языковыми моделями, типа GPT-х, показывают, что и они не способны решить задачу по планиметрии олимпиадного уровня [1].

Предлагаемый метод сочетает в себе лучшее из обоих направлений. Идея подхода в том, чтобы на основе символического языка генерировать большое количество синтетических задач с решениями для обучения языковой модели. В статье описывается структура и компоненты DNA, а также обсуждаются модификации и потенциал применения в контексте других задач.

Структура и ключевые особенности

Дедуктивно-нейронный подход предполагает наличие трёх компонент – дедуктивного, языкового и доменного. Последний примечателен тем, что его элементы (рис.1) зависят от области решаемой задачи. Он не универсален в отличие от других двух.



Рис.1. Схема доменного компонента

Центральным звеном выступает символический язык, описывающий понятия и аксиоматику в предикатной

форме. Например, факт того, что M – середина отрезка AB , выразится в предикат $mdr(M, A, B)$ [2].

Рассмотрим *дедуктивный движок (DE)* (рис.2). Первый его компонент, *дедуктивный граф G* , отвечает за хранение предикатов и организацию логических связей между ними. Предыдущие работы использовали концепцию дедуктивного датафрейма [2] или deductive database [4].

Отметим, что G является ориентированным, ациклическим гиперграфом, вершинами которого являются предикаты. Если предикат S является предпосылкой для предиката P , то S и P соединяются ребром. Для каждой вершины графа необходимо уметь решать задачу о *поиске минимального обоснования*:

$$\forall P, \exists \bar{G}(P) \in G: \bar{G}(P) \rightarrow P, |\bar{G}(P)| \rightarrow \min \quad (1)$$



Рис.2. Устройство дедуктивного движка

Представление G как гиперграфа решает проблему транзитивности, коллинеарности и цикличности. *Группа* предикатов с транзитивными свойствами представлены в виде подграфов (рис.3). Алгоритмом поиска в ширину ищется кратчайший путь в подграфе (иными словами, минимальный набор предикатов) для определения истинности элемента группы.

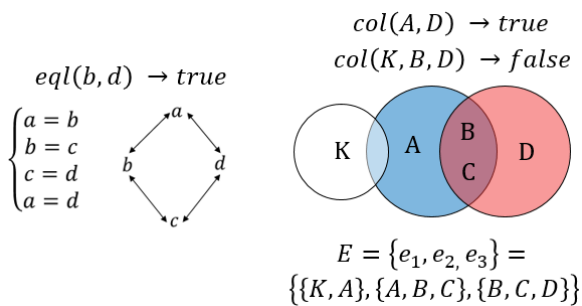


Рис.3. Подграфы транзитивности

Для k -транзитивности задача в общем случае сводится к поиску минимального остовного дерева [1] жадным алгоритмом.

Назначение *логического семплера условий* в генерации решаемой задачи – условие и вопрос в предикатной форме. Кортеж $(\{P_i\}, C, \bar{G}(C))$ – задание и решение соответственно, сериализуется в текстовый

формат для добавления в *синтетический набор данных* языковой модели (рис.4).

Вычислительный модуль (NM) представляет предикаты, выражающие количественные факты (например, $\frac{AB}{AC} = \frac{HC}{AM}$), в виде системы линейных уравнений (СЛУ). Решая СЛУ методом Гаусса, можно получить не только численные, но и логические факты (например, $(NM) x = y \rightarrow eql(x, y)$ (DE)). Вычислительный и дедуктивный модули применяются поочередно.

Для решения задачи поиска минимального обоснования целевого уравнения с коэффициентами b используется метод целочисленного линейного программирования (ЛП).

Пусть M – количество уравнений, N – переменных. Дана матрица коэффициентов A . Определяем целые неотрицательные векторы $x, y \in R^M$ и решаем задачу ЛП:

$$x, y = \min_{x,y} \sum_i (x_i + y_i): A^T(x - y) = b \quad (2)$$

Получаем, что i -е уравнения (i -е строки в A), для таких i , что $x_i - y_i \neq 0$, образуют минимальную СЛУ, из которой можно вывести целевое уравнение [1].

Генератор нестандартных решений языковой модели формирует список предложений на естественном языке вида ‘Построй точки X такие, что ...’.

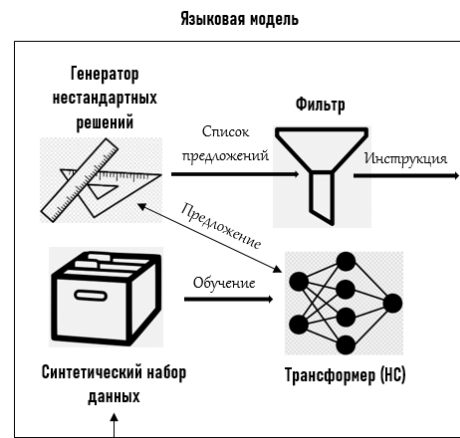


Рис.4. Языковая модель

Фильтр с помощью алгоритма лучевого поиска (beam search [1]) выбирает первые k элементов из списка, предложенного генератором.

Можно выделить следующие преимущества подхода:

Генерация нестандартных решений. Главной трудностью для чисто дедуктивного подхода выступают дополнительные построения (ДП). Их невозможно перебрать и можно лишь запрограммировать в рамках эвристики [3]. Однако DNA на основе языковой модели способен предоставлять список построений на естественном языке. Фильтр (рис.4) ранжирует ДП по некоторому критерию [1], выбирает из них первые k и переводит на символичный язык. Предла-

гается ввести определение ДП (нестандартного шага), как пару $\{x, P(x)\}$, где x – это новый объект, а $P(x)$ – предикат, описывающую логическую связь добавленного с имеющимися объектами.

Заметим, что механизм создания нестандартного решения легко обобщается и не привязан к геометрии. Например, если решаются задачи по комбинаторике, то предложение использовать метод раскраски (ввести чёрные и белые клетки) является примером «дополнительного построения».

Процедура сбора и разметки данных не нужна. Нейронная сеть обучается на сколь угодно большом количестве искусственно сгенерированных данных. Подчеркнём, что все данные – корректные, так как их синтезом занимается дедуктивный модуль, который гарантирует правильность полученного решения.

Открытие новых закономерностей. В результате обучения DNA получается набор искусственно сгенерированных доказанных теорем, и как показывает исследование [1], из них известна лишь малая часть. Более того, модель из [1] обнаружила в задании IMO (International Mathematical Olympiad) лишнее условие. Следовательно, у DNA есть потенциал к обобщению уже известных фактов.

Для каждой области нужен свой символичный язык. Формализация любой научной сферы потребует отдельной разработки собственного символического языка и теорем. Безусловно, это главный недостаток DNA. Возможно, в будущем удастся установить общие принципы организации этого процесса.

Модификации и результаты

Результатом работы является DNASolver – новый сервис, основанный на AlphaGeometry [1], который реализует дедуктивно-нейронный подход. Рассмотрим пример их работы.

Для обеих моделей решение выглядит следующим образом:

* From theorem premises:

A B C D: Points
 $BD \perp AC$ [00]
 $CD \perp AB$ [01]

* Auxiliary Constructions:

E: Points
 B, E, D are collinear [02]
 A, E, C are collinear [03]

* Proof steps:

001. B, E, D are collinear [02] & A, E, C are collinear [03] & $BD \perp AC$ [00] $\Rightarrow \angle BEA = \angle CED$ [04]

...

008. $BE:EC = AE:ED$ [09] & $\angle AED = \angle BEC$ [05] (Similar Triangles) $\Rightarrow \angle EAD = \angle EBC$ [11]

009. $\angle ADE = \angle BCE$ [10] & B, E, D are collinear [02] & A, E, C are collinear [03] & $\angle EAD = \angle EBC$ [11] $\Rightarrow AD \perp BC$

В DNASolver можно вводить условие как на естественном языке, так и на предикатном (рис.5):

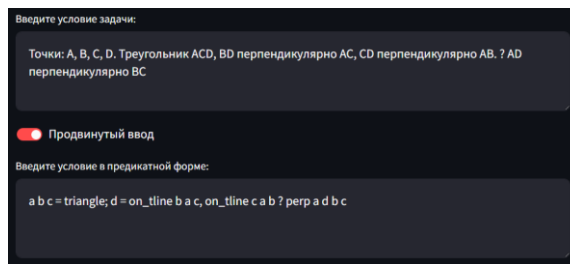


Рис.5. Условие задачи, DNASolver

Сервис формирует запрос к ChatGPT [5], что нужно перевести задачу, основываясь на сформированном в ходе работы небольшом (~35) наборе данных [6], содержащем условие задач IMO с их переводом на предикатный язык.

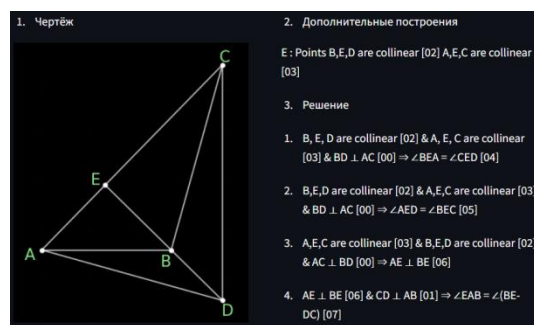


Рис.6. Вывод решения, DNASolver

В дедуктивный движок из [1] были внесены следующие изменения – программно реализованы комплексные теоремы, не входящие в школьную программу (например, теорема Рейма). Проведённое исследование показывает, что тем шире спектр известных теорем (аксиоматика), тем лучше метрика качества ДП (отношение удачных ДП к общему числу предложенных) (см. Таблицу 1).

Таблица 1. Оценки для различных подходов

| Подход | Верные реш. | Метрика ДП |
|---------------|-------------|------------|
| DE+NM | 21 | - |
| AlphaGeometry | 25 | 82,8% |
| DNASolver | 25 | 85,9% |
| GPT-4 | 0 | - |

Заключение

В работе сформулированы составляющие дедуктивно-нейронного подхода, установлены его преимущества и потенциал применения в других областях. Приведены и проанализированы результаты работы системы DNASolver.

Список литературы

- Trinh, T.H., Wu, Y., Le, Q.V. et al. Solving olympiad geometry without human demonstrations. Nature 625, 476–482 (2024). <https://doi.org/10.1038/s41586-023-06747-5>
- Просвирнин Р.О., Применение предикатного подхода в автоматизации процесса решения задач планиметрии. / Мой вклад в науку: сборник статей Международного конкурса молодых учёных. - МЦНС «Наука и Просвещение», 2023. – с. 8-12.

3. Geometric Automated Theorem Proving: электронный журнал. - URL: <https://www.mat.uc.pt/~pedro/cientificos/presentationUrbino2019.pdf>. (дата обращения: 09.02.2024). - Текст: электронный.

4. Chou, S. C., Gao, X. S. and Zhang, J. Z.: Automated generation of readable proofs with geometric invariants, I: Multiple and shortest proof generation, J. Automated Reasoning 17(3) (1996), p 325–347.

5. -URL: <https://openai.com/blog/chatgpt> (дата обращения: 12.02.2024). - Текст: электронный.

6. -URL: <https://github.com/enlacroix/neurogeometry/blob/main/IMOproblems.xlsx> (дата обращения: 12.02.2024). - Текст: электронный.

МОДУЛЬ ВИЗУАЛЬНОГО ПРОГРАММИРОВАНИЯ ДЛЯ РЕДАКТОРА МАТЕРИАЛОВ В ИГРОВЫХ ДВИЖКАХ

*Ковалев Р.А.¹, Ляпунова С.А.²,
Маркин А.А.², Попов Л.Д.²*

Национальный исследовательский университет

«Высшая школа экономики»,

¹департамент прикладной математики,

²департамент компьютерной инженерии

МИЭМ НИУ ВШЭ

Аннотация

В исследовании подробно рассматриваются вопросы применения методов визуального программирования в редакторе материалов. В центре внимания - практическое использование технологии на языке C#, включающее в себя работу с шейдерами для выполнения специфических функций обработки изображений, таких как поворот, размытие и отзеркаливание. Эти функции были выбраны за их значимость и частое использование в обработке графического контента. Визуальное программирование может значительно упростить и ускорить разработку, предоставляя разработчикам интуитивно понятные инструменты для более эффективной работы.

Введение

В сфере программирования специалисты сталкиваются с необходимостью применения определенных алгоритмов, понимание и освоение которых может представлять сложность даже при наличии исходного кода. Данная проблематика актуальна для IT-компаний, где часто наблюдается ситуация, когда разработанный младшим специалистом (junior) алгоритм подвергается доработке более опытным коллегой (senior). Эта проблема обусловлена различиями в уровне знаний и навыков сотрудников, а также их пониманием задачи.

В этой связи был произведен обзор аналогов систем визуального программирования, включая такие известные платформы как Scratch [1], Blockly [2], Node-RED [3], LabVIEW [4], MakeCode [5] и Blueprints [6]. Из обзора можно сделать вывод, визуальное программирование актуальное направление в сфере разработки.

В области разработки компьютерных игр дизайнерам и разработчикам приходится создавать уникальные визуальные элементы для текстурирования объектов, такие как изменение времен года, отражение в зеркалах и другие специфические эффекты. Традиционные методы часто влекут за собой значительные временные затраты на поиск или создание необходимых ресурсов, причем разработанные решения могут быть использованы ограниченным числом специалистов.

В этом контексте визуальное программирование представляет собой перспективное решение, способное облегчить и оптимизировать процесс разработки. Целесообразность разработки комплексной системы визуального программирования, основанной на принципах редактора материалов, обусловлена потребностью в унифицированных и гибких инструментах для работы с текстурами изображений. Такая система должна включать в себя блоки (ноды) с настраиваемыми параметрами, позволяя пользователям эффективно и интуитивно управлять процессом создания и модификации визуального контента.

Архитектура разрабатываемой системы

В рамках разрабатываемой системы выделены три основных типа блоков: функциональный, константный и блок триггера. Каждый из блоков обладает несколькими слотами, которые могут быть классифицированы как путевые или слоты данных, в зависимости от их функционального назначения.

Блоки

В представленной системе визуального программирования ключевым элементом являются блоки, выполняющие функцию основных структурных компонентов. Блоки предоставляют пользователю разнообразный набор действий и функций, которые могут быть комбинированы в произвольном порядке, обеспечивая тем самым расширение функциональных возможностей разрабатываемой программы. В рамках разработки были созданы четыре типа блоков:

1. Функциональный – блок, который предназначен для выполнения конкретных функций, заранее определенных разработчиком.

2. Триггер – блок, который имеет событие-триггер, инициирующее процесс обхода графа методом обхода в глубину.

3. Константный – этот блок предназначен исключительно для хранения информации, определенной пользователем.

4. Блоки шейдеров – тип блоков, предназначенный для работы с шейдерами.

Каждый блок в системе содержит входные и/или выходные слоты, а также переменные, используемые в соответствующих функциях. Отдельное внимание уделяется возможности редактирования этих переменных пользователем, что позволяет оптимизировать процесс настройки и повышает понятность программы. Такой подход способствует уменьшению количества входных параметров функций и снижает необходимость в большом числе константных блоков, что в свою очередь минимизирует визуальную

нагрузку и упрощает восприятие структуры программы.

Применение блоков в качестве основных структурных элементов системы визуального программирования таким образом способствует созданию гибкой и интуитивно понятной среды разработки, позволяя пользователям эффективно реализовывать сложные алгоритмы и функциональные возможности.

Слоты

В рамках разработки системы визуального программирования особое внимание уделяется структуре слотов, которые представляют собой ключевые элементы каждого блока. В исследовании выделены два основных типа слотов: слоты данных и путевые слоты. Слоты данных функционируют как аргументы и возвращаемые значения функций, тогда как путевые слоты определяют последовательность выполнения программы.

Каждый слот должен содержать информацию о типе хранимого значения и перечне связанных с ним слотов. В случае путевых слотов тип хранимого значения представляет собой нулевой указатель, что указывает на отсутствие прямой ассоциации с конкретными данными. Напротив, выходные слоты предназначены для хранения данных в соответствии с определенным типом.

Такая структурированность и классификация слотов обеспечивает гибкость и точность в определении логики работы блоков в системе визуального программирования. Данная архитектура слотов способствует эффективной реализации функциональных задач и упорядоченному управлению потоками данных и управления в процессе программирования.

Шейдеры

В контексте предстоящей работы по обработке изображений важным техническим решением является применение шейдеров [7-8]. Основным преимуществом использования шейдеров является их способность исполняться на процессорах видеокарт (GPU), что предоставляет значительные выгоды как для разработчиков, так и для пользователей.

Одной из ключевых характеристик шейдеров является их эффективность в обработке графических данных, обеспечиваемая за счет специализированных вычислительных ресурсов GPU [8-9]. Это позволяет существенно разгрузить центральный процессор (CPU) компьютера, предотвращая его перегрузку и связанные с этим риски, такие как перегрев или потенциальные поломки. В результате использование GPU для обработки изображений повышает общую производительность системы и обеспечивает более стабильную и безопасную работу.

Таким образом, интеграция шейдеров в систему обработки изображений является стратегически важным решением, направленным на оптимизацию вычислительных процессов и повышение надежности работы программного обеспечения

Блоки шейдеров

В рамках разработки редактора материалов была осуществлена реализация трех функциональных блоков, предназначенных для выполнения следующих операций над изображениями: размытие по Гауссу, поворот и отзеркаливание. Методы, использованные для реализации этих блоков, подробно описаны ниже.

Размытие по Гауссу

Размытие изображения по Гауссу [9] является методом обработки, направленным на сглаживание изображения и уменьшение уровня шума. Этот метод основан на применении функции распределения Гаусса к значениям пикселей. В контексте попиксельной обработки, каждый пиксель изображения подвергается воздействию весов, определенных Гауссовым распределением, что приводит к усреднению значений в окрестности каждого пикселя.

Поворот изображения

Поворот изображения на определенное количество градусов включает в себя изменение угла направления каждого пикселя на определенный угол в плоскости изображения. Это создает впечатление вращения всего изображения вокруг своего центра.

Отзеркаливание изображения

Отзеркаливание изображения – это операция, изменяющая порядок пикселей изображения, создавая зеркальное отражение вдоль вертикальной оси. В контексте попиксельной обработки, каждый пиксель изображения заменяется пикселем, симметричным относительно выбранной оси.

Результаты использования функциональных блоков

С использованием описанных методов были реализованы разнообразные функциональные блоки шейдеров. В частности, созданы следующие блоки: Blur, Mirror effect и Rotation. Каждый из этих блоков выполняет обработку входного изображения, возвращая результат с соответствующим эффектом. Например, блок Blur осуществляет размытие по Гауссу с учетом заданного радиуса, а Mirror effect создает отраженное изображение относительно оси OY. Rotation () позволяет поворачивать фотографию на заданный угол. Блок представлен на рис.1.

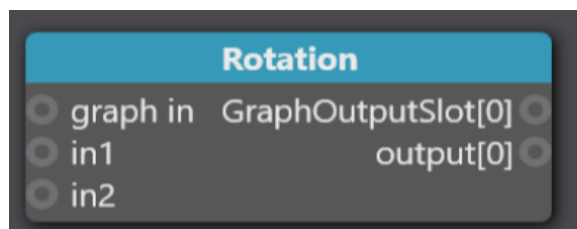


Рис.1. Пример блока-шейдеров Rotation

Дополнительно к указанным эффектам, в реализации присутствуют два дополнительных блока: Image и Save. Блок Image предоставляет путь к фото-

графии без каких-либо входных параметров. С другой стороны, блок Save сохраняет переданное изображение, принимая на вход путь и само изображение. Введен блок-триггер Button entry point, который предоставляет возможность инициировать выполнение процесса.

Благодаря данной системе с архитектурой визуального программирования и блоками шейдеров, существует возможность строить графы любого функционала.

Каждый блок-шейдер может быть использован индивидуально, а также возможно их объединение в более обширные последовательности для реализации комплексных визуальных эффектов. Данная функциональность обеспечивает возможность модификации изображений путём последовательного применения множества шейдеров.

Для проверки совместимости и интеграции данных блоков были объединены все доступные шейдеры в единую последовательность. В итоге изображение подверглось вертикальному отражению, последующему размытию и, в завершение, повороту. Результаты тестирования системы представлены на рис.2.

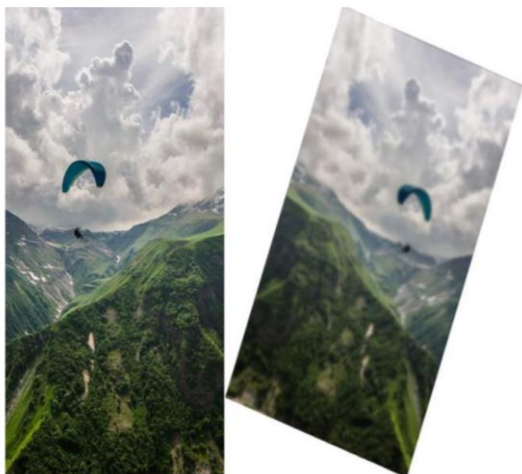


Рис.2. Результат работы системы

Заключение

На основе полученных результатов можно сделать вывод, что модуль визуального программирования способствует ускорению процесса разработки программ и повышает их наглядность для пользователя. В рамках проведенной работы были реализованы следующие компоненты для системы редактора материалов: три основных блока для обработки изображений, один блок-триггер для инициации алгоритма обхода графа, а также два вспомогательных блока – константный блок, хранящий целочисленное значение, и блок для загрузки изображений из файла.

Список литературы

1. Maloney, J., Resnick, M., Rusk, N., Silverman, B., & Eastmond, E. (2010). The Scratch Programming Language and Environment. *ACM Transactions on Computing Education*, 10(4), Article 16, 1-15. <https://dl.acm.org/doi/abs/10.1145/1868358.1868363>

2. Prasetyo Adi, P. D., & Kitagawa, A. (2020). A Review of the Blockly Programming on M5Stack Board and MQTT Based for Programming Education. *IEEE Xplore*, 1-13. <https://ieeexplore.ieee.org/abstract/document/8994922>

3. Rajalakshmi, A., & Shahnasser, H. (2018). Internet of Things using Node-Red and Alexa. *IEEE Xplore*, 1-16.

<https://ieeexplore.ieee.org/abstract/document/8261194>

4. Bitter, R., Mohiuddin, T., & Nawrocki, M. (2017). *LabView: Advanced Programming Techniques*. pp. 1-69. <https://books.google.ru/books?id=L-fKBQAAQBAJ>

5. Ball, T., Chatra, A., de Halleux, P., Hodges, S., Moskal, M., & Russell, J. (2019). Microsoft MakeCode: embedded programming for education, in blocks and TypeScript. *ACM Digital Library*, 7-12. <https://dl.acm.org/doi/abs/10.1145/3358711.3361630>

6. Romero, M., & Sewell, B. (2022). *Blueprints Visual Scripting for Unreal Engine 5*. pp. 1-21. <https://books.google.ru/books?id=yhZuEAAAQBAJ>

7. Tornai, R., & Fürjes-Benke, P. (2020). Compute Shader in Image Processing Development. *CEUR Workshop Proceedings*, November 6–8, 2020. <https://ceur-ws.org/Vol-2874/short20.pdf>

8. Persson, G., & Udd, J. (2010). Ray Tracing on GPU: Performance comparison between the CPU and the Compute Shader with DirectX 11. pp. 30-50. <https://www.divaportal.org/smash/record.jsf?pid=diva2%3A829795&dsid=1642>

9. Brodtkorb, A. R., Hagen, T. R., & Sætra, M. L. (2013). Graphics processing unit (GPU) programming strategies and trends in GPU computing. *Journal of Parallel and Distributed Computing*, 73(1), 1-15. <https://doi.org/10.1016/j.jpdc.2012.04.003>

10. Bozkurt, F., Yağanoğlu, M., & Günay, F. B. (2015). Effective Gaussian Blurring Process on Graphics Processing Unit with CUDA. *International Journal of Machine Learning and Computing*, 5(1), 57-61. <https://www.ijml.org/vol5/483-W012.pdf>

ПРОЦЕДУРНАЯ ГЕНЕРАЦИЯ ТРЕХМЕРНОЙ МЕСТНОСТИ В ВИРТУАЛЬНОМ ПРОСТРАНСТВЕ

Макшанчиков В.С.¹, Романов И.А.², Казаков Л.К.¹, Матанов К.А.³

Национальный исследовательский университет «Высшая школа экономики»,

¹ департамент прикладной математики,

² департамент электронной инженерии,

*³ департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В данной работе исследованы методы качественного и эффективного построения трехмерной местности в виртуальном пространстве с использованием алгоритмов процедурной генерации.

Был разработан функционал создания и изменения рельефа, его разделения на биомы и последую-

щего насыщения игровыми объектами для процедурной генерации трехмерного ландшафта с применением генератора когерентного шума и разбиения его на зоны в соответствии с определенными характеристиками. Пользователь имеет возможность редактировать сгенерированный рельеф с помощью внутренней системы коррекции. Для расстановки игровых элементов в созданной местности был применён алгоритм расстановки игровых объектов вдоль вершин графа, построенного для ландшафта.

Введение

В современных областях, связанных с компьютерной графикой, существует постоянная потребность в создании трехмерных визуальных ассетов. Для создания сложных игровых объектов привлекают 3D-художников, которые вручную разрабатывают многополигональные модели. Однако в случаях, когда требуется создание большого объема контента, использование 3D-художников становится неэффективным. В таких ситуациях используются алгоритмы процедурной генерации, позволяющие создавать контент на основе заранее заданных алгоритмов, минимизируя ручной труд [1].

Однако ситуации, при которых пользователь полностью удовлетворён результатом процедурной генерации, достаточно редки. Таким образом, чтобы избежать многократной регенерации модели местности, необходимо предоставить пользователю возможность вносить изменения в результат работы алгоритмов вручную, тем самым позволив ему создавать свои уникальные решения при построении ландшафта.

Существует несколько инструментов для решения схожей проблемы: Gaia, World Machine, World Creator, Houdini и Unity: Gaia. Вышеупомянутые инструменты являются внешними модулями, которые предоставляют функционал в виде отдельных интерфейсов внутри среды разработки компьютерных игр или в отдельных приложениях. Предлагаемое решение будет встроенным в саму среду разработки компьютерных игр, что позволит пользователю использовать интегрированный функционал, учитывающий архитектуру среды разработки компьютерных игр для процедурной генерации мира и его дальнейшей настройки.

Перечисленные инструменты не предоставляют пользователю возможности развить генерацию ландшафта в генерацию полноценного виртуального окружения. В частности ни один из них не поддерживает полноценное создание точек интереса, являющихся ключевым аспектом для создания осмысленно наполненного мира, то есть содержащего объекты, являющиеся результатом уникальных событий до момента непосредственного его наблюдения.

Подсистема генерации базовой формы ландшафта

Для получения разнородного ландшафта был использован генератор когерентного шума, позволяющего наслаиваются друг на друга шумы, в ходе чего генерируется карта высот местности. Когерентный

шум - это вид гладкого псевдослучайного шума. Данный шум создается когерентной функцией шума, которая обладает тремя важными свойствами:

1. Подача одного и того же входного значения всегда приведет к получению одного и того же выходного значения.

2. Небольшое изменение входного значения приведет к небольшому изменению выходного значения.

3. Большое изменение входного значения приведет к случайному изменению выходного значения.

Алгоритм основан на работе с трехмерными когерентными шумами, принимающими на вход несколько параметров:

- Лакуарность - мера неоднородности заполнения шумом пространства

- Октавы - количество накладываемых друг на друга функций шума

- Частота - характеристика периодичности первой октавы шума

- Стойкость - математическая характеристика, определяющая соотношение частоты и амплитуда

Для получения разнообразных форм ландшафта было реализовано 4 вида шумов: Шум Перлина (Perlin - англ.), Вздвигающийся шум (Billow - англ.), Ребристый шум (Ridged - англ.), Смешанный шум (совокупность Billow, Perlin) [2 - 3].

Алгоритм генерации базовой формы с использованием шумов, наслаивающихся друг на друга с уменьшающейся амплитудой и увеличивающейся частотой, генерирует карту высот, представленную на рис. 1. На её основе строится меш [4]. Итогом работы подсистемы является генерация ландшафта трёхмерной местности.

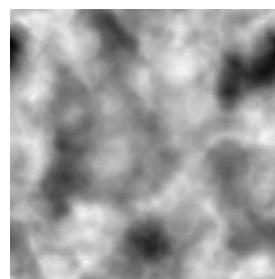


Рис.1. Карта высот

Подсистема метаданных

Данная подсистема начинает работу после завершения формирования ландшафта подсистемой базовой формы. Она принимает на вход следующие параметры:

- Количество полигонов

- Количественная характеристика рек и озер как внутриигровых объектов наполнения ландшафта

- Настройки разделения игрового мира на биомы

- Под биомом подразумевается совокупность полигонов, обладающая уникальными отличительными параметрами и свойствами.

Настройки разделения игрового мира на биомы подразумевает задание пользователем системы характеристик, а именно высоты, влажности, обеспе-

чивающих однозначное определение принадлежности полигона к биому.

Поверхность ландшафта разделяется на области малого по отношению к размеру всего ландшафта размера. Размер области косвенно задаётся пользователем через входной параметр количества полигонов и рассчитывается внутренним алгоритмом в зависимости от размера игрового мира (входной параметр подсистемы базовой формы). Каждая из этих областей в соответствии со своим расположением, а также на основе входных данных, приобщена к одному из биомов, т.е. набору из таких областей.

Выделение биомов реализовано с помощью комбинации графа Воронного [5] и триангуляции Делоне [6 - 7]. Структура областей имеет полигональный вид в силу использования вершин и рёбер диаграммы Вороного в том же смысле (точки 1 и 2 на рис.2). Центрами данных полигонов являются вершины триангуляции Делоне, соседство областей определяется ребрами триангуляции (вершина А и В на рис.2).

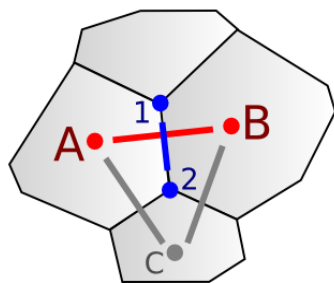


Рис.2. Полигонализация посредством комбинации графа Воронного и триангуляции Делоне

Получив разделение ландшафта на биомы, внутренний алгоритм генерирует двумерную текстуру, где каждый полигон окрашен в заданный пользователем цвет и принадлежит соответствующему биому. Результатом выполнения алгоритма является текстура, представленная на рис.3 и адаптированная под уже сгенерированный предыдущей подсистемой ландшафт с разделением его на биомы и уже отображёнными проекциями статических объектов в двумерном пространстве.

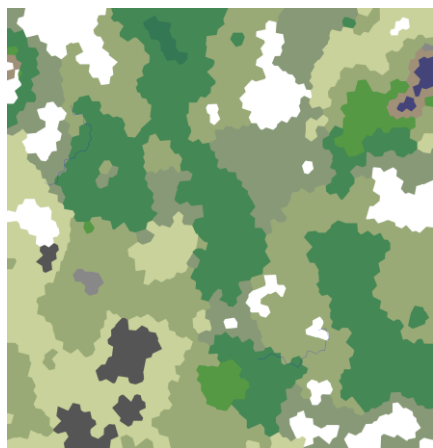


Рис.3. Карта метаданных

Подсистема ручной коррекции

Для достижения максимального контроля над конечным результатом генерации был разработан функционал ручной коррекции базовой формы ландшафта. Данный инструмент позволяет изменять форму рельефа с помощью кисти, определяемой следующими параметрами:

- Форма кисти
- Тип кисти (понижение, повышение или выравнивание)
- Вращение кисти
- Размер кисти
- Отступ
- Сила кисти

Некоторые из указанных параметров имеют дополнительные настройки, с помощью которых возможно указать рамки, в которых они будут случайным образом изменять свое значение для достижения эффекта реалистичного ландшафта.

Точка, в которой применяется ручная коррекция, определяется путем проекции двумерного пространства экрана пользователя на трехмерное пространство террейна с помощью соответствующей матрицы [8]. В случае, если расстояние от данной точки до предыдущей меньше введенного отступа, изменение ландшафта не производится.

При применении кисти к точке происходит изменение формы кисти в соответствии с текущими параметрами вращения и размера с учетом их случайной составляющей. Полученная форма определяет вершины меша, которые необходимо сместить, а также коэффициент, применяемый к силе кисти для каждой вершины меша, определяя числовое значение смещения вершин (см. рис.4). Сила кисти в данном случае определяется пользователем в виде числового значения [9].

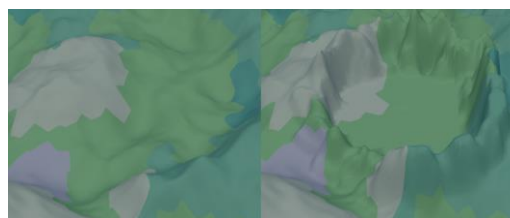


Рис.4. Пример изменения рельефа при ручной коррекции

Подсистема процедурной расстановки игровых объектов

Получив желаемую геометрию ландшафта [10] и распределение террейна на биомы, пользователь получает возможность наполнить местность игровыми объектами окружения. Подсистема начинает свою работу после всех предыдущих и принимает на вход следующие параметры:

- Ассеты и их количество
- Биомы нахождения как отдельных ассетов, так и точек интереса. Точка интереса (Point of interest) - внутриигровая область, которая представляет определенный интерес для пользователей, с набором игровых объектов.

• Категория важности (для ассетов, принадлежащим точкам интереса)

Игровые объекты, переданные пользователем, устанавливаются в центрах сгенерированных ранее полигонов. Окружающие объекты (деревья, камни и т.п.) расставляются в рамках указанного биома в случайном порядке, при этом случайно меняется их размер и угол поворота для создания более разнообразного наполнения игрового мира, в то время как расстановка объектов точек интереса происходит сложнее.

Для генерации точки интереса определяется центр биома с помощью алгоритма, который проходит по сгенерированному в подсистеме метаданных полигонам, определяя расстояния между точками. Он проходит от одной из точек, находящейся на краю биома, по всем полигонам данного биома, определяя расстояние от стартового полигона до всех остальных, после чего выбираются полигоны со средним значением расстояния и повторяются те же действия среди полученного набора точек, таким образом определяется центр биома. От найденного центра начинается расстановка моделей точки интереса, радиально отдаляясь от него, при этом учитываемые определенные пользователем категории важности: сначала устанавливаются модели первой категории, потом - второй, затем третьей, самыми последними используются модели четвертой категории. Данное разбиение объектов позволяет создавать более реалистичные города, в центре которых стоят особые здания, а на краю более типичные постройки. Объекты точек интереса также имеют случайный угол поворота и размер.

Результаты

Конечным результатом работы является последовательное объединение всех перечисленных подсистем (см. рис. 5).

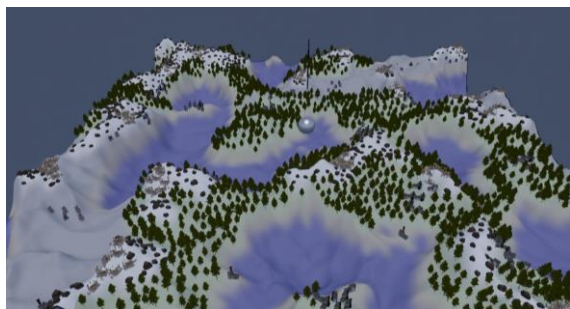


Рис.5. Конечный результат работы всех подсистем

Заключение

В ходе исследования было рассмотрено несколько инструментов, позволяющих генерировать трёхмерный мир. На основе выявленных недостатков был предложен новый алгоритм для моделирования виртуального мира, позволяющий поэтапно генерировать ландшафт, создавать произвольные биомы, расставлять объекты окружения и точки интереса и вручную редактировать полученный результат.

Список литературы

1. Short T., Adams T. (ed.). Procedural generation in game design. – CRC Press, 2017.
2. Lagae A. et al. State of the Art in Procedural Noise Functions //Eurographics (State of the Art Reports). – 2010. – С. 1-19.
3. Stephenson I. Noise //Essential RenderMan®. – 2007. – С. 201-212.
4. Sorkine O. Differential representations for mesh processing //Computer Graphics Forum. – Oxford, UK : Blackwell Publishing Ltd, 2006. – Т. 25. – №. 4. – С. 789-807.
5. Aurenhammer F. Voronoi diagrams—a survey of a fundamental geometric data structure //ACM Computing Surveys (CSUR). – 1991. – Т. 23. – №. 3. – С. 345-405.
6. Chen L., Xu J. Optimal delaunay triangulations //Journal of Computational Mathematics. – 2004. – С. 299-308.
7. Lee D. T., Schachter B. J. Two algorithms for constructing a Delaunay triangulation //International Journal of Computer & Information Sciences. – 1980. – Т. 9. – №. 3. – С. 219-242.
8. E. Penelope Holland, James N Aegerter, Calvin Dytham, Graham C Smith // Landscape as a Model: The Importance of Geometry - URL: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0030200#references> (дата обращения: 19.02.2024).
9. Lacroix, P.; Levoy, M. Fast volume rendering using a shear-warp factorization of the viewing transformation // In Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques – Orlando, FL, USA, 1994 – С. 451–458
10. Dong Wang, Yunan Zhang, Peng Tian, Nanming Yan Real-Time GPUbased Visualization of Tile Tracks in Dynamic Terrain – 2009

СИНТЕЗ РЕЧИ С ЗАДАНЫМИ ПАРАМЕТРАМИ

Сиротинский Н.В., Горбунов И.И., Дусов М.А., Яковлев М.В., Якубов В.Ю.
*Национальный исследовательский университет «Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В данной работе рассматривается задача реализации нейросетевого алгоритма синтеза человеческой речи из текста с сохранением особенностей произношения целевого объекта. В представленном тексте описывается процесс сбора и подготовки обучающих данных, а также обучения нейросетевой модели Tacotron 2 в связке с вокодером HiFiGAN и GST эмбедингами, применяемыми для стилизации генерируемой речи.

Введение

Задачу синтеза речи нельзя назвать новой, первые разработки в сторону ее современного решения по-

явились еще в прошлом веке (важной вехой можно назвать создание вокодера в 1930-х годах), а современные технологии позволяют генерировать речь практически не отличимую от настоящей. Уже сейчас такие системы применяются для озвучивания перевода в онлайн-переводчиках, как элементы интерфейса для слабовидящих и слепых людей, в некоторых голосовых ассистентах. С развитием технологии можно ожидать и расширение области применения синтеза речи из текста (text-to-speech, TTS), начиная от автоматического озвучивания игр и заканчивая генерацией речи на различных языках с различными акцентами для лингвистов и людей, изучающих язык.

Что касается технологии переноса голоса и синтеза речи с заданными параметрами, то актуальность исследования можно обусловить отставанием от существующих методов генерации обманных образов на изображениях и, в частности, видеозаписях. Так называемые deepfake получили большое распространение в последнее время, при этом существующие методы клонирования голоса несколько отстают в развитии. Немаловажным моментом в исследовании переноса голоса является необходимость распознавания факта подмены голоса, поскольку данная технология может быть использована для различных противоправных действий. Исследование особенностей обработки и реалистичной генерации и клонирования голоса в перспективе позволят также улучшить методы распознавания.

Сбор обучающих данных

Предлагаемое в данной работе решение базируется на методах глубокого обучения. Применяемым нейросетевым алгоритмам для обучения требуется большое количество данных. Таким образом, часть работы посвящена сбору русскоязычного набора данных из открытых источников.

Обучающие данные состояли из пар аудиодорожек и их транскрипций, полученных с 3 видеохостингов Youtube, VK видео и Rutube. Для перечисленных сервисов вручную собирались списки URL'ов, которые впоследствии обкачивались и транскрибировались отдельной моделью. В приоритете был сбор аудиокниг и подкастов, без постороннего шума.

Для транскрибирования аудио на предложения была выбрана нейросетевая модель Whisper [1]. Whisper – это предобученная модель, основанная на архитектуре энкодер-декодер и предназначенная для распознавания речи. В данной работе для транскрипции данных использовалась дополнительно дообученная на русскоязычных аудио версия модели Whisper-small [1] – данная модель является оптимальной с точки зрения точности и скорости. Во время транскрипции аудиофайлов, модель генерирует временные метки для каждого слова, что дает возможность разделить исходный аудиофайл на множество аудиофайлов, для каждого фрагмента которого известна транскрипция.

Всего описанным образом было собрано около 1000 часов озвученного текста. После обработки бы-

ло оставлено порядка 443 часов речи. Основные параметры полученных данных приведены в таблице 1.

Таблица 1. Характеристики наборов данных для обучения

| | Датасет вариативной речи одного спикера | Многоголосый датасет |
|-----------------------------|---|----------------------|
| Суммарная длина аудио, часы | 53 | 443 |
| Число уникальных слов | 100081 | 373824 |
| Минимальная длина аудио, с | 2 | 2 |
| Максимальная длина аудио | 8 | 8 |
| Минимальная длина текста | 4 | 11 |
| Максимальная длина текста | 100 | 100 |
| Максимальное число слов | 24 | 25 |
| Количество пар | 44471 | 365510 |

Обработка звуковых данных

Звуковые данные представляют собой последовательность значений амплитуды и одним из ключевых параметров звукового сигнала является частота дискретизации, обычно принимающая значения от 16 кГц до 48 кГц. Обработать такие длинные последовательности сложно, поэтому часто используется преобразование в частотное представление или спектрограмму с помощью дискретного преобразования Фурье. Мел-шкала [2] позволяет выделить значимую информацию о звуке, уменьшая объем данных и ускоряя обработку. Преобразования звука в мел-спектрограмму являются необратимыми, и для восстановления звука используются классические методы, такие как алгоритм обращения мел-масштаба и итеративный алгоритм Граффин-Лима [3], а также более продвинутые нейросетевые подходы, которые восстанавливают аудио более качественно, например, HifiGAN [4] или WaveGlow [5].

Генерация речи с заданными параметрами

Акустические модели занимаются преобразованием текста в аудио и являются базовым элементом большинства TTS систем. Одной из популярных акустических моделей является Tacotron [6], который использует нейросетевые алгоритмы для преобразования текста в речь. Tacotron состоит из нескольких компонентов: первый компонент извлекает признаки из текста, второй генерирует линейную спектрограмму на основе этих признаков, и третий компонент – алгоритм Граффин-Лима для восстановления аудиозаписи из спектрограммы.

С целью улучшения качества генерации речи была разработана модель Tacotron2 [7]. Она представляет собой улучшенную версию оригинального Tacotron. Главной особенностью Tacotron2 является использование измененного механизма внимания — local sensitive attention [8]. Это позволяет устранять артефакты заикания и повторения, присущие Tacotron. Также в модели присутствует отдельный модуль предсказания остановки генерации аудио, что позволяет модели работать без дополнительной постобработки и удаления тишины в конце синтезированной речи. Благодаря тому, что Tacotron2 генерирует mel-спектрограммы, в отличие от своего предшественника, он может работать с более качественными вокодерами, что дополнительно повышает качество синтезированной речи.

Таким образом, Tacotron2 является более продвинутой и эффективной моделью для генерации речи, чем оригинальный Tacotron, и обладает возможностью создавать более естественные и высококачественные аудиозаписи.

Тем не менее оба вышеописанных алгоритма не могут применяться в задаче генерации аудио с заданными параметрами, например, генерации аудио, с определенным эмоциональным окрасом или же генерации речи с заданным тембром голоса. Для того, чтобы интегрировать такой функционал в модель мы использовали модуль GST [9]. Данный модуль при инициализации модели создает набор токенов, который не меняется в процессе обучения и используется для сравнения с входным аудио для выделения из него основных особенностей и характеристик в виде вектора, который при объединении с признаками текста будет влиять на стиль генерируемого аудио. Данный модуль реализует в своей архитектуре подход self-attention, который используется в различных современных моделях трансформеров.

В оригинальной статье GST используется вместе с архитектурой tacotron, но в данной работе было решено интегрировать его в Tacotron2. Это является одним из нововведений предлагаемой реализации, которое должно улучшить качество генерируемой речи. В данной работе итоговый вектор из GST конкатенируется с выходным вектором энкодера Tacotron2 и имеет такую же размерность в 256 элементов.

Еще одним улучшением оригинальной архитектуры является измененный процесс обучения Tacotron2, который в нашем подходе генерирует не 1 фрагмент mel-спектрограммы, а сразу 3. Такой подход называется reduction фактор при обучении, который сильно сокращает объем вычислений и увеличивает скорость сходимости модели, сохраняя при этом почти оригинальное качество генерации.

Восстановление аудио из mel-представления

Для восстановления аудио из mel-спектрограмм применяют вокодеры. В данной работе используется нейросетевая вокодер HifiGAN [4], особенностью которого является сверточная генеративно-состязательная архитектура (GAN) [10].

Гиперпараметры входных мел-спектрограмм использовались такие же, как и в акустической модели.

При обучении использовались фрагменты аудио длиной 8192, что соответствует длине мел-спектрограммы 32. Таким образом размерность входных данных при обучении составляла 80×32 , а выходных - 1×8192 . При синтезе модель принимает мел-спектрограммы произвольной длины.

Для обучения вокодера использовался тот же набор обучающих данных, дополнительно очищенный с помощью предобученного Denoiser [11] — модели, которая позволяет очищать аудио от фоновых шумов, выделяя речь. Это является важным аспектом, поскольку собранные в открытых источниках данные содержат много лишних звуков помимо речи, а вокодеры, обученные на таких данных, склонны переносить эти шумы в синтезируемую речь, снижая ее качество.

Дообучение модели

Акустическая модель и вокодер обучались отдельно, несмотря на то что обе модели работают с mel-спектрограммами, при их совместной работе возникают различные артефакты, например, дрожание голоса. Для того, чтобы это исправить, необходимо их совместно дообучить. В данной работе было решено дообучать именно вокодер, что заключается в дообучении уже обученной модели на mel-спектрограммах, предсказанных акустической моделью, в данном случае tacotron2. Это позволяет вокодеру подстроиться под конкретное распределение и артефакты акустической модели, исправляя их в сторону реальных данных и получая речь высокого качества.

Результаты

Tacotron2 обучался более 150 тысяч итераций с размером батча 64, при этом использовались параметры mel-преобразования, описанные в оригинальной статье [7]. HifiGAN обучался с теми же параметрами mel-преобразования 1.2 миллиона итераций и размером батча 16.

Для дообучения сначала были получены mel-спектрограммы для каждого аудиофайла и его текста с помощью Tacotron2. Далее, используя эти mel-спектрограммы в качестве входных данных, HifiGAN дообучался в течение 60 тысяч итераций и размером батча 16.

В результате был получен алгоритм приведенный на рис.1.

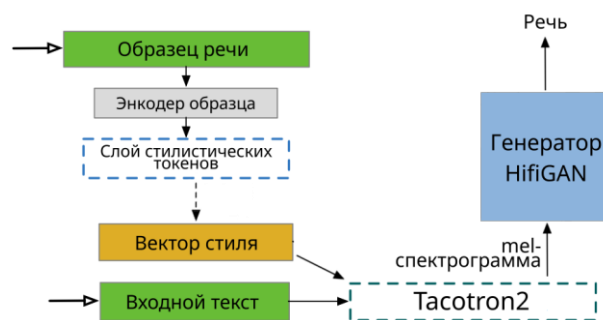


Рис.1. Схема нейросетевого алгоритма синтеза речи с заданными параметрами

Для оценки качества работы генеративных моделей обычно применяется метрика средней оценки мнения или mean opinion score (MOS), представляющая собой опрос. В данной работе был использован предобученный нейросетевой аналог данной метрики, способный на основании образца речи предсказывать оценку MOS, близкую к реальным опросам [12]. Таким образом был выбран пример реальной речи как образец для оценки, после чего для каждого текста и аудио из валидационной выборки был синтезирован искусственный пример речи и вычислено для него значение метрики MOS. В результате были получены усредненные значения метрики, представленные в таблице 1.

Таблица 1

Усредненные значения метрики MOS для алгоритма синтеза речи

| | MOS |
|----------------|------|
| Ground Truth | 4.27 |
| Без дообучения | 3.39 |
| С дообучением | 4.06 |

Как видно из результатов, дообучение вокодера дало сильный прирост качества синтезируемой речи.

Заключение

В данной работе был описан реализованный нейросетевой алгоритм синтеза речи из текста с заданным голосом. Много внимания было уделено сбору, разметке и очистке данных из открытых источников, для чего применялись две преобученные нейросети — Whisper и Denoiser. В качестве акустической модели для синтеза речи применялся Tacotron2 с модулем GST для выделения глобального стиля и особенностей голоса целевой речи. Для восстановления полученных mel-спектрограмм применялся вокодер HiFiGAN. Акустическая модель и вокодер были дополнительно дообучены для улучшения качества их совместной работы.

Список литературы

1. Radford A. et al. Robust speech recognition via large-scale weak supervision //International Conference on Machine Learning. – PMLR, 2023. – С. 28492-28518.
2. Combrinck H., Botha E. C. On the mel-scaled cepstrum //department of Electrical and Electronic Engineering, University of Pretoria. – 1996.
3. Griffin D., Lim J. Signal estimation from modified short-time Fourier transform //IEEE Transactions on acoustics, speech, and signal processing. – 1984. – Т. 32. – №. 2. – С. 236-243.
4. Kong J., Kim J., Bae J. HiFi-gan: Generative adversarial networks for efficient and high fidelity speech synthesis //Advances in Neural Information Processing Systems. – 2020. – Т. 33. – С. 17022-17033.
5. Prenger R., Valle R., Catanzaro B. Waveglow: A flow-based generative network for speech synthesis //ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2019. – С. 3617-3621.

6. Wang Y. et al. Tacotron: Towards end-to-end speech synthesis //arXiv preprint arXiv:1703.10135. – 2017.

7. Shen J. et al. Natural tts synthesis by conditioning wavenet on mel spectrogram predictions //2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). – IEEE, 2018. – С. 4779-4783.

8. Bahdanau D., Cho K., Bengio Y. Neural machine translation by jointly learning to align and translate //arXiv preprint arXiv:1409.0473. – 2014.

9. Wang Y. et al. Style tokens: Unsupervised style modeling, control and transfer in end-to-end speech synthesis //International conference on machine learning. – PMLR, 2018. – С. 5180-5189.

10. Goodfellow I. et al. Generative adversarial nets //Advances in neural information processing systems. – 2014. – Т. 27.

11. Defossez A., Synnaeve G., Adi Y. Real time speech enhancement in the waveform domain //arXiv preprint arXiv:2006.12847. – 2020.

12. Manocha P., Kumar A. Speech quality assessment through MOS using non-matching references //arXiv preprint arXiv:2206.12285. – 2022.

РАЗРАБОТКА ДЕТЕКТОРА ПЕЧАТИ НА ИЗОБРАЖЕНИЯХ

Прокудина К.В., Скрипленок М.С.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Исследована эффективность различных методов и алгоритмов для создания детектора печатей на изображениях. Создана собственная модель машинного обучения, которая была дообучена на новых данных. Проведено экспериментальное сравнение различных моделей YOLOv8 и сделаны выводы об их эффективности.

Введение

В настоящее время в области компьютерного зрения и обработки изображений одной из актуальных задач является разработка детектора печатей на документах. Подобная задача [0] имеет широкое применение в различных сферах, таких как банковское дело, юридическая сфера, медицина, документооборот, контроль подлинности документов и другие. Для обучения нейронной сети необходимо иметь набор размеченных данных, содержащих изображения документов с печатями и без них. После получения приемлемой оценки качества модели можно применять детектор для автоматического обнаружения печатей на новых документах, при этом он будет выделять только те области, в которых есть различные штампы на картинках.

Выявление проблемы

Причина, по которой необходимо разработать свой детектор печати на изображениях, заключается в том, что существующие методы имеют некоторые ограничения и недостатки. Например, один из алгоритмов [2], основанный на HLS-сегментации [3] и преобразовании Хафа [4], работает только с цветными изображениями и зависит от размера документа. Он также может срабатывать на рукописи и шумы, что может привести к ложным срабатываниям. Другой подход [5] объединяет детектирование объектов и их сопоставление, но имеет проблемы с архитектурой и библиотеками, а также недочеты в алгоритмической схеме, влияющие на метрики машинного обучения. Еще в статье [6] авторы исследуют обнаружение и сегментацию штампов на отсканированных документах, но проблема усложняется различными размерами, формами, сложностями и цветами чернил штампов, а также качеством отсканированных документов и сложным фоном. В работе [7] исследователи предлагают новый метод сегментации и классификации штампов на основе цветовой кластеризации и геометрических признаков, но такой подход требует дальнейшей оптимизации и корректировки. Наконец, в статье [8] разработан новый подход к обнаружению штампов на изображениях с помощью компактной и эффективной модели нейронной сети, основанной на детекторе YOLOv3 [9]. Однако, есть несколько недостатков: ограниченный объем данных, оценка производительности модели только на одном типе изображений, отсутствие расширения функциональности для классификации различных типов штампов и отсутствие экспериментов в реальных условиях.

Таким образом, разработка собственного детектора печатей на изображениях становится актуальной задачей, чтобы преодолеть ограничения и недостатки существующих методов. Необходимо разработать алгоритм, который будет работать не только с цветными картинками, но и с различными размерами и формами штампов. Также важно учитывать сложности, связанные с качеством и фоном документов. Алгоритм должен использовать современные техники обработки изображений, такие как нейронные сети, сегментацию и классификацию на основе различных признаков.

Работа с данными и разметка

Опираясь на статьи, описывающие данную тему [2, 5-8], сначала собираются данные для обучения модели. Важно отметить, что использование печатей направлено на подтверждение подлинности документа, особенно подписи должностного лица. Чтобы определить печати, необходимо понять, какие типы печатей существуют. В данном случае выделяются три типа печатей: гербовые, эквивалентные гербовым и обычные. Главная задача – это нахождение любых печатей в документе независимо от того, к какой организации они относятся.

Парсинг сайтов (сбор данных путем извлечения информации с различных веб-сайтов), браузерное расширение (процесс сбора данных, просматривая

различные веб-страницы и извлекая нужную информацию), пок кадровая нарезка видео (нарезка видео на отдельные кадры), где есть необходимый и интересный объект, – все это является первостепенным и важным этапом, который позволяет получить набор изображений с печатями для дальнейшей обработки и анализа. В процессе работы было собрано около 1000 изображений. Это может быть выполнено путем поиска и скачивания изображений из различных источников, таких как интернет или базы данных. Следующим шагом является разметка данных. С помощью универсального инструмента LabelImg [10] необходимо нанести ограничивающие рамки (bbox) [10] на все объекты, содержащие атрибуты, такие как форма, цвет, надписи печатными буквами, герб и т.д., что позволит понять, это штамп/печать или нет.

Дообучение модели

Когда данные собраны, можно проводить эксперименты и смотреть, какая модель [11] лучше детектирует печати на изображениях.

В первую очередь, надо рассмотреть YOLOv8 [12], поскольку у нее высокая скорость обработки, точность детекции, метрики, такие как полнота (recall) [13], точность (precision) [13], и средний показатель точности (mean average precision, mAP) [14].

Для детектирования объектов на изображениях сначала проводится обучение модели YOLOv8n [12] на заданном датасете, включающем изображения печатей разных форм, видов и размеров, к тому же с добавлением аугментации: поворот относительно центра, отражение, обрезка, размытие. В задаче детектируется только один класс, поэтому данных вполне достаточно для дообучения. Количество эпох (итераций обучения) составляет 20, так как определяется единственная категория и изображений относительно немного, большее число эпох может вызвать переобучение. Размер входных картинок равен 640x640 (стандартное значение по умолчанию) с использованием билинейной интерполяции, значения пикселей нормализуются. Были попытки обучить модели на протяжении большего числа эпох, но выигрыша в точности не получено. После этого берется обученная модель для детектирования объектов на тестовых изображениях (148 шт.) с определенным порогом уверенности (threshold = 0.1) и анализируются полученные предсказания (рис.1).



Рис. 1. Предсказание модели на валидационной выборке

При применении различных версий Yolo8 (Yolo8n, Yolo8s, Yolo8m) [12] для детектирования печатей, можно заметить значительные различия в скорости обработки изображений. Предобработка включает загрузку изображения в память, изменение размера и нормализацию. Постобработка включает декодирование вывода, NMS и пересчёт координат ограничивающих прямоугольников. Время потерь - время вычисления потерь. Предобработка и постобработка необходимы, чтобы получить предсказания, получив на вход изображения. Время вычисления потерь не влияет на качество модели, так как имеет смысл только во время обучения. Например, при обучении модели Yolo8n время обработки изображения составляет всего 0.2 миллисекунды на предварительную обработку, 3.2 миллисекунды на вывод, 0.0 миллисекунды на потери и 1.3 миллисекунды на постобработку для каждого изображения. Сравнивая это с другими версиями, такими как Yolo8s и Yolo8m, можно увидеть значительные различия. Например, для модели Yolo8s время предварительной обработки составляет 0.5 миллисекунды, время вывода - 6.7 миллисекунды, время потерь - 0.0 миллисекунды и время постобработки - 0.7 миллисекунды на каждое изображение. А для модели Yolo8m соответствующие значения составляют 0.6 миллисекунды, 7.6 миллисекунды, 0.0 миллисекунды и 0.6 миллисекунды.

Таким образом, можно заметить, что модель Yolo8n обеспечивает значительно более высокую скорость обработки изображений по сравнению с другими версиями Yolo8 [12], что делает ее более эффективной для детектирования печатей, особенно в режиме реального времени.

Получены также количественные характеристики: ошибка в позиции и размерах прямоугольника (Bounding Box Regression Loss) - 0.379, ошибка в значении вероятности класса (Classification Loss) - 0.317. Эти две метрики помогают определить, насколько хорошо модель может найти местоположение объектов на изображении и правильно классифицировать объекты.

Дальше находятся максимальное значение метрики metrics/mAP50-95(B) [15] и соответствующую ему эпоху обучения. По спискам labels [10], precision, recall [13] и mAP50 [14], содержащим соответствующие значения метрик для каждой модели и эпохи с максимальными значениями, создается график для отображения столбчатых диаграмм метрик для каждой модели, при этом mAP50 [14] представляет собой среднюю точность обнаружения объектов при различных уровнях отсечения (threshold) для вероятности обнаружения объекта, является важной метрикой для детектирования печатей, поскольку она учитывает как точность, так и полноту обнаружения объектов, что позволяет оценить качество работы модели на практике.

Как видно по диаграмме (рис.2), на которой отображаются важные метрики, сложно определить различие, между моделями практически нет никакого отличия, однако надо не забывать, что nano гораздо быстрее и еще выше точность даёт.

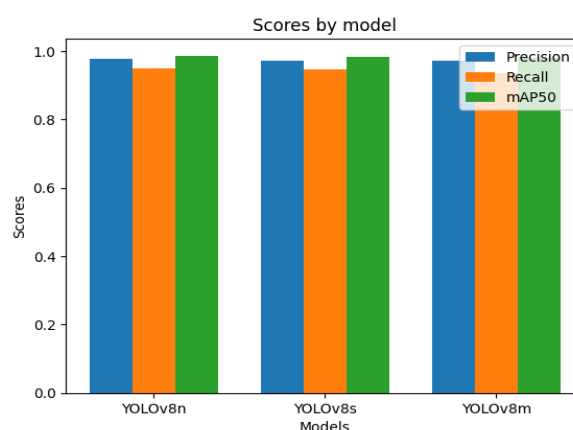


Рис.2. Анализ типов моделей

Однако, разница в процентах для точности между yolo8n и yolo8s составляет: [0.2994%], а для yolo8n и yolo8m [0.2339%]. Nano показывает лучший результат по всем параметрам: производительности, эффективности, точности, полноте и средней точности модели при различных порогах перекрытия (IoU) между предсказанными и истинными объектами. В Таблице 2 приведен анализ трех моделей по наиболее важным метрикам: precision, recall [13] и mAP50 [14]. Но важно не забывать, что последний параметр [14] дает финальное решение.

Таблица 2. Анализ моделей

| Модели | Метрики | | |
|---------|-----------|---------|---------|
| | precision | recall | mAP50 |
| YOLOv8n | 0.97574 | 0.94876 | 0.98626 |
| YOLOv8s | 0.97281 | 0.94633 | 0.98365 |
| YOLOv8m | 0.97251 | 0.93551 | 0.98037 |

Для наглядности приведены графики (рис.3) с предсказаниями модели YOLOv8n - это Mean Average Precision (средняя средняя точность, или mAP) [14]. В случае с YOLOv8n [12], mAP вычисляется при пороге уверенности (confidence threshold) равном 0.5, поэтому он обозначается как mAP@0.5 или mAP50 и является расчетной метрикой, которая оценивает качество обнаружения объектов на изображениях. Использование mAP50 в качестве главной метрики для детектирования печатей с помощью YOLOv8n [12] позволяет оценить точность детектирования и обнаружения печатей на изображениях. Чем выше значение mAP50 [14], тем лучше модель справляется с задачей детектирования печатей. 0,98626 – максимальный показатель mAP50.

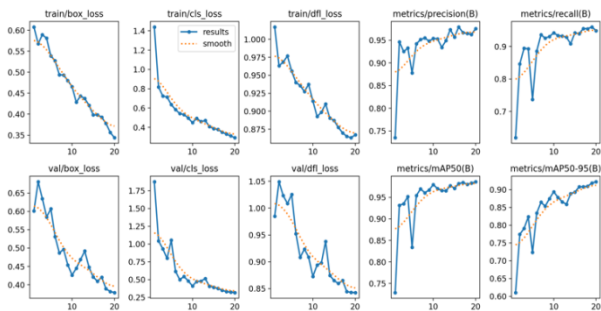


Рис.3. Показатели модели во время обучения

Дальше планируется сделать более быструю модель, которая не будет проигрывать в точности. Еще можно добавить функционал сравнения найденных печатей с эталонными.

Заключение

В ходе анализа проведен обзор существующих решений для детектирования печатей на изображениях, который выявил все ограничения и недостатки. Для решения этой проблемы собран и размечен датасет на 1000 картинок с различными типами штампов. Проведенное исследование показало, что модель YOLOv8 "nano" обладает высокой точностью и производительностью при детектировании печатей на изображениях. Кроме того, была определена важная метрика - mAP50, которая учитывает как precision, так и recall обнаружения объектов и позволяет оценить качество работы модели на практике.

Стоит отметить, работа будет еще усовершенствоваться: планируется добавить сравнение печатей с эталонными и опробовать различные для них методы. В будущем планируется создать модель, которая не только превзойдет все существующие аналоги, но и обеспечит высокую точность в работе, и будет быстрее всех рассмотренных ранее аналогов.

Список литературы

1. How to Detect Objects in Images Using the YOLOv8 Neural Network [Электронный ресурс]. – URL: <https://www.freecodecamp.org/news/how-to-detect-objects-in-images-using-yolov8/> (дата обращения 10.01.2024).
2. Распознавание печатей: нейронные сети против SIFT, и причем тут Госзакупки [Электронный ресурс]. – URL: <https://habr.com/ru/companies/solarsecurity/articles/580488/> (дата обращения 10.01.2024).
3. HLS-сегментация [Электронный ресурс]. – URL: <https://cqr.company/ru/wiki/protocols/understanding-http-live-streaming-hls-the-future-of-online-video-delivery/> (дата обращения 11.01.2024).
4. Преобразование Хафа. – URL: <https://habr.com/ru/articles/102948/#:~:text=Преобразование%20Хафа%20—%20это%20метод%20 обнаружения, изображения%20множества%20кривых%20 заданного%20семейства> (дата обращения 11.01.2024).

5. Реализация поиска печатей на OpenCV без нейронок, регистрации и смс. – URL: <https://habr.com/ru/companies/directum/articles/486476/> (дата обращения 11.01.2024).

6. Stamp Detection using Computer Vision and Python [Электронный ресурс]. – URL: <https://medium.com/analytics-vidhya/stamp-detection-using-computer-vision-and-python-6038a0243429> (дата обращения 11.01.2024).

7. Stamp Detection in Color Document Images [Электронный ресурс]. – URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065485> (дата обращения 11.01.2024).

8. Fast and Accurate Deep Learning Model for Stamps Detection for Embedded Devices [Электронный ресурс]. – URL: <https://link.springer.com/article/10.1134/S1054661822040046> (дата обращения 11.01.2024).

9. YOLOv3 [Электронный ресурс]. – URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKewiM4s-SxPGDaxX_PhA1HfjIDMoQFnoECAoQAQ&url=https%3A%2F%2Fdocs.ultralytics.com%2Fru%2Fmodels%2Fyolov3%2F&usq=AOvVaw3uVizHH874oxsSdhYojw3D&opi=89978449 (дата обращения 11.01.2024).

10. LabelImg [Электронный ресурс]. – URL: <https://github.com/HumanSignal/labelImg> (дата обращения 11.01.2024).

11. Object detection guide for Python [Электронный ресурс]. – URL: https://developers.google.com/mediapipe/solutions/vision/object_detector/python (дата обращения 15.01.2024).

12. Ultralytics YOLOv8 [Электронный ресурс]. – URL: <https://docs.ultralytics.com/ru/> (дата обращения 15.01.2024).

13. Precision and recall. Как они соотносятся с порогом принятия решения [Электронный ресурс]. – URL: <https://habr.com/ru/articles/661119/> (дата обращения 17.01.2024).

14. MAP50 в детекции объектов [Электронный ресурс]. – URL: <https://ml.i-neti.ru/map-mean-average-precision/> (дата обращения 17.01.2024).

15. Difference between precision and map50:95 [Электронный ресурс]. – URL: <https://github.com/ultralytics/yolov5/issues/8039> (дата обращения 17.01.2024).

АВТОМАТИЗАЦИЯ ПОСТРОЕНИЯ МОДЕЛИ ДЕФОРМАЦИОННОГО ПОВЕДЕНИЯ СВЕРХПЛАСТИЧНЫХ МАТЕРИАЛОВ

Анохин В.В., Миколаенко В.В., Кудряшов М.Д.
 Национальный исследовательский университет
 «Высшая школа экономики»,
 департамент прикладной математики
 МИЭМ НИУ ВШЭ

Аннотация

Сверхпластическая газовая формовка – это один из способов обработки материалов давлением, используемый при производстве тонкостенных изделий сложной геометрической формы. Для проектирования технологических режимов производства таких изделий используется компьютерное моделирование технологического процесса, для которого необходимо задать свойства материала. Основной экспериментальный метод исследования свойств сверхпластичных материалов – одноосное растяжение плоских образцов с постоянной скоростью деформации. Однако, стандартная процедура интерпретации экспериментальных данных таких испытаний не учитывает особенности течения материала в образце, что приводит к ошибкам при построении модели деформационного поведения материала. В данной работе предлагается алгоритм коррекции такой модели, учитывающий неоднородность течения материала в образце. Для автоматизации процедуры построения свойств материала было разработано соответствующее программное обеспечение, на вход которого принимаются экспериментальные данные механических испытаний образцов на одноосное растяжение.

Введение

Сверхпластичность – это состояние материала, в котором он способен накапливать большие деформации при бездефектном формоизменении [1]. Для проектирования технологического процесса сверхпластической газовой формовки используется компьютерное моделирование, для которого необходимо знать модель деформационного поведения материала [2]. Основным экспериментальным методом исследования свойств материала является испытание плоских образцов на одноосное растяжение с постоянной скоростью деформации [3]. Процедура проведения таких испытаний и интерпретация экспериментальных данных описана в международном стандарте ASTM-E2448 [4]. В указанной процедуре интерпретации экспериментальных данных пренебрегается неоднородностью распределения скоростей деформации в образце. Ввиду высокой чувствительности материала в сверхпластичном состоянии к скорости деформации, такое пренебрежение приводит к ошибкам при построении модели деформационного поведения материала порядка 20% [5].

В данной работе предлагается итерационная процедура коррекции свойств материала. Для того чтобы учесть неоднородность распределения скоростей деформации в образце, на каждой итерации проводится конечно-элементное моделирование рассматриваемых испытаний.

Испытания образцов на растяжение

Экспериментальными данными механических испытаний плоских образцов на растяжение с постоянной скоростью деформации является набор значений силы сопротивления образца захватам F и длины рабочей зоны l в разные моменты времени. Исходя из гипотезы о том, что растяжение в рассматриваемом механическом испытании является идеально одноосным, полученные значения далее используются при расчёте деформации ε и напряжения σ [4]:

$$\varepsilon = \left| \ln \left(\frac{l}{l_0} \right) \right|, \quad (1)$$

$$\sigma = \frac{F l}{V_0}, \quad (2)$$

где l_0 и V_0 – исходные длина рабочей зоны и объём образца соответственно. Полученный набор значений ε и σ представляет из себя кривую «напряжение-деформация».

В рассматриваемых испытаниях поддерживается постоянная скорость деформации $\dot{\varepsilon}_{trg}$, за счет непрерывного возрастания скорости перемещения захватов v :

$$v = l \dot{\varepsilon}_{trg}. \quad (3)$$

В данной работе для получения экспериментальных данных были проведены испытания плоских образцов толщиной 1 мм. Геометрия образца и установки изображена на рис. 1.

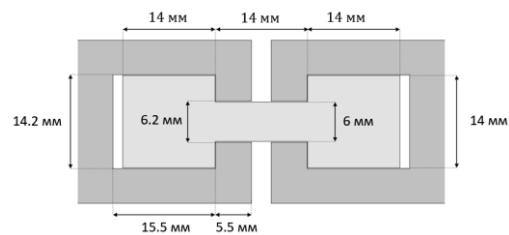


Рис. 1. Геометрия образца и инструмента

Кривые «напряжение-деформация», полученные из разных испытаний, используются при построении модели деформационного поведения материала в следующем разделе.

Построение модели материала

Модель деформационного поведения материала в общем виде представляет из себя зависимость напряжения от деформации и скорости деформации:

$$\sigma = \sigma(\varepsilon, \dot{\varepsilon}) \quad (4)$$

Единственная кривая «напряжения-деформации» не позволяет построить модель материала, описываемую зависимостью (4), так как при проведении испытания скорость деформации постоянна. Проведение серии испытаний с одной и той же геометрией, но различными скоростями деформации $\dot{\varepsilon}_{trg}$ позволяет решить данную проблему. Построение модели материала заключается аппроксимации набора кривых «напряжения-деформации», полученного по результатам серии испытаний.

Зависимость напряжения от скорости деформации в логарифмических осях выражается с помощью сигмоидной функции. В узком диапазоне скоростей деформации, данная функция аппроксимируется уравнением Бэкофена [6]:

$$\sigma(\varepsilon, \dot{\varepsilon}) = k\dot{\varepsilon}^m \quad (5)$$

В данной работе коэффициенты k и m выражаются в виде полиномиальной зависимости от деформаций следующим образом:

$$k(\varepsilon) = K_0 + K_1\varepsilon + \dots + K_5\varepsilon^5, \quad (6)$$

$$m(\varepsilon) = M_0 + M_1\varepsilon + \dots + M_5\varepsilon^5. \quad (7)$$

Конечно-элементное моделирование

Процедура, описанная в прошлом разделе, использует формулы (1-2) при построении набора кривых «напряжения-деформации». В данных формулах пренебрегается неоднородностью течения материала в образце, что приводит к ошибкам в расчёте кривых «напряжение-деформация». Компьютерное моделирование испытаний образцов на растяжение, позволяет учесть особенности течения материала в рассматриваемом образце и по выходу моделирования скорректировать исходные кривые.

В данной работе было реализовано программное обеспечение, позволяющее проводить такое моделирование с помощью метода конечных элементов. Моделирование проводится для 1/8 части образца, так как остальные части деформируются симметрично. В качестве конечных элементов использовались призмы с треугольным основанием, которое лежит в плоскости симметрии, а ребра всегда перпендикулярны основанию. Такой вид элемента позволяет снизить число степеней свободы с 18 для произвольной призмы до 9.

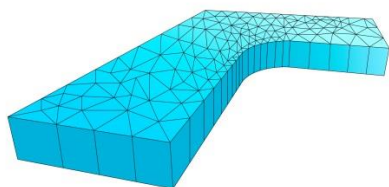


Рис.2. Разбиение образца на призмы

Выбор таких конечных элементов позволяет учесть изменение толщины образца, пренебрегая сдвиговыми деформациями, влияние которых на результаты моделирования незначительно. При этом

время расчёта моделирования становится сравнимым с двумерным моделированием.

На рис.3 показано распределение интенсивности скорости деформации при проведении испытания на одноосное растяжение с $\dot{\varepsilon}_{trg} = 0.01c^{-1}$, которое удаётся учесть, с помощью разработанного программного обеспечения.

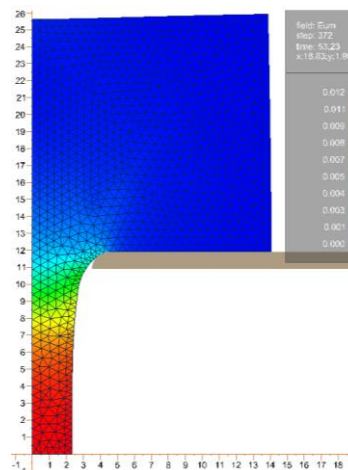


Рис.3. Распределение интенсивности скорости деформации в образце

По результатам компьютерного моделирования, аналогично реальному испытанию, определяется сила F и длина l , по которым далее, с помощью уравнений (1-2) рассчитываются кривые «напряжение-деформация». Такие кривые используются в процедуре коррекции при сравнении выхода моделирования с экспериментальными данными, что описано в следующем разделе.

Коррекция модели материала

Процедура коррекции свойств материала была предложена в работе [7]. Рассматриваемая процедура является итерационной. На вход i -ой итерации поступают кривые «напряжение-деформация» $\sigma_{in}^{[i]}$, на основе которых рассчитываются свойства материала. Далее эти свойства используются при компьютерном моделировании серии испытаний, по результатам которых проводится построение кривых «напряжение-деформация» $\sigma_{out}^{[i]}$, аналогично тому, как это было сделано для экспериментальных данных. Кривые «напряжение-деформация», подаваемые на вход следующей итерации $\sigma_{in}^{[i+1]}$, рассчитываются исходя из разницы между выходом моделирования $\sigma_{out}^{[i]}$ и кривыми σ_{ref} , построенными по экспериментальным данным:

$$\sigma_{in}^{[i+1]} = \frac{\sigma_{ref}}{\sigma_{out}^{[i]}} \sigma_{in}^{[i]}(\varepsilon, \dot{\varepsilon}), \quad (8)$$

где $\sigma_{in}^{[i]}(\varepsilon, \dot{\varepsilon})$ – свойства материала, используемые при конечно-элементном моделировании на i -ой итерации.

На вход нулевой итерации поступают кривые $\sigma_{in}^{[0]} = \sigma_{ref}$. Критерием выхода из итерационной процедуры является:

$$\Delta_i = \left\| \sigma_{out}^{[i]} - \sigma_{ref} \right\| < \delta, \quad (9)$$

где $\|\cdot\|$ – среднеквадратичная норма, δ – минимальное допустимое значение. Выполнение условия (9) в силу формул (1,2) свидетельствует о приближении силы F и длины l , полученной при моделировании, к соответствующим экспериментальным данным. Следовательно, истинная модель материала $\sigma_{in}^{[i]}(\varepsilon, \dot{\varepsilon})$ построена с заданной точностью δ .

Результаты

Описанная в прошлом разделе процедура коррекции свойств материала в данной работе была апробирована на сплаве системы Al-Mg-Fe-Ni [8]. При сотрудничестве с лабораторией «Ультрамелкозернистых металлических материалов» НИТУ «МИСиС» были получены экспериментальные данные для серии механических испытаний, проведённые со следующими скоростями деформации: $\dot{\varepsilon}_{trg} = 0.06c^{-1}$, $0.03c^{-1}$, $0.01c^{-1}$, $0.005c^{-1}$, $0.002c^{-1}$.

На рис.4 точками обозначены полученные экспериментальные данные в виде кривых «напряжение-деформация». Сплошной линией обозначены кривые «напряжение-деформация», полученные по результатам первых трех итераций процедур коррекции.

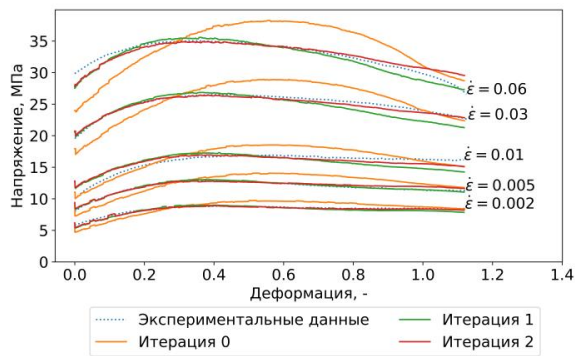


Рис.4. Сравнение выходов моделирования с экспериментальными данными

На рис.5 представлена эволюция моделей свойств материала, используемых на первых трёх итерациях.

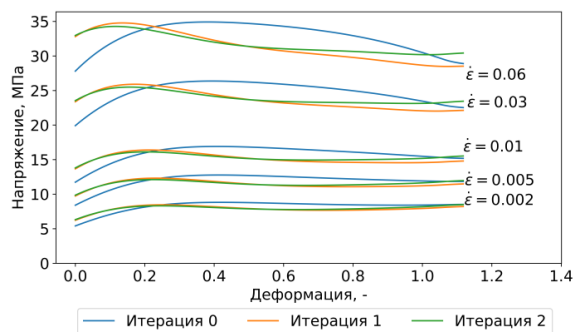


Рис.5. Эволюция модели материала

Всего было проведено 4 итерации процедуры коррекции. Среднеквадратичная норма разности между σ_{ref} и $\sigma_{out}^{[i]}$ менялась следующим образом: $\Delta_0 \approx 1.83$, $\Delta_1 \approx 0.79$, $\Delta_2 \approx 0.57$, $\Delta_3 \approx 0.57$.

Разница между среднеквадратичными нормами разности, соответствующими выходам моделирования на 2 и 3 итерации, пренебрежимо мала. Поэтому итерационная процедура завершается на 3 итерации с результатом в виде скорректированной модели. Параметры такой модели представлены в таблице 1.

Таблица 1. Параметры модели

| | $i = 0$ | $i = 1$ | $i = 2$ | $i = 3$ | $i = 4$ | $i = 5$ |
|-------|---------|---------|---------|---------|---------|---------|
| K_i | 128.25 | -114.16 | 41.34 | 356.32 | -551.23 | 228.29 |
| M_i | 0.48 | -0.59 | 1.55 | -1.73 | 0.74 | -0.07 |

Заключение

В данной работе рассмотрен алгоритм коррекции свойств материала, построенных по стандартной процедуре интерпретации экспериментальных данных испытаний на одноосное растяжение плоских образцов. Коррекция проводится исходя из результатов компьютерного моделирования, позволяющих учесть неоднородность распределения скоростей деформации в образце. Для автоматизации предложенной процедуры коррекции было разработано соответствующее программное обеспечение.

Алгоритм коррекции был опробован на экспериментальных данных, полученных из механических испытаний образцов сплава системы Al-Mg-Fe-Ni. Полученная модель материала была представлена в виде уравнения Бэкофена, в которой коэффициенты k и m выражаются в виде полиномиальной зависимости от деформаций ε . Ошибка построенной модели материала удалось уменьшить с $\Delta = 1.83$ для исходной модели до $\Delta = 0.57$ для скорректированной модели. Ошибка вычислялась согласно формуле (9).

Список литературы

1. Sherby, O. D. Superplasticity-Recent advances and future directions / O. D. Sherby, J. Wadsworth. – 1989.
2. Deshmukh, P. V. Study of superplastic forming process using finite element analysis / P. V. Deshmukh. – 2003.
3. Loveday, M. S. Tensile testing of metallic materials: A review / M. S. Loveday, T. Gray, J. Aegerter // Final report of the TENSTAND project of work package. – 2004. – Vol. 1.
4. E2448-11. Standard Test Method for Determining the Superplastic Properties of Metallic Sheet Materials / E2448-11 // ASTM Book of Standards. – 2011.
5. Aksenov, S. The effect of material properties on the accuracy of superplastic tensile test / S. Aksenov, V. Mikolaenko // Metals. – 2020.
6. Enikeev, F. U. An analysis of the superplastic forming of a thin circular diaphragm / F. U. Enikeev, A. A. Kruglov // International Journal of Mechanical Sciences. – 1995.
7. Миколаенко, В. Коррекция модели деформационного поведения сверхпластических

материалов / В. Миколаенко // Материалы межвузовской научно-технической конференции студентов, аспирантов и молодых специалистов имени Е.В. Арменского. – 2022. – Р. 34-37.

8. Al-Mg-Fe-Ni based alloy for high strain rate superplastic forming / A. A. Kishchik, A. V. Mikhaylovskaya, A. D. Kotov [et al.] // Materials Science and Engineering: A. – 2018. – Vol. 718. – P. 190-197.

СТРУКТУРА ДАННЫХ ДЛЯ ВЫДЕЛЕНИЯ КЛАСТЕРОВ ЧАСТИЦ В МОЛЕКУЛЯРНОМ МОДЕЛИРОВАНИИ

Вараксин И.М., Писарев В.В.

*Национальный исследовательский университет «Высшая школа экономики»,
международная лаборатория
суперкомпьютерного атомистического
моделирования и многомасштабного анализа*

Аннотация

Работа посвящена реализации структуры данных, позволяющей хранить и обрабатывать данные о кластерах точек в пространстве. Структура применяется для выделения кластеров атомов в молекулярном моделировании. Программный интерфейс для структуры реализован авторами на языке Julia в рамках библиотеки анализа данных молекулярного моделирования MDPProcessing.jl. Описаны структуры данных, позволяющие оптимизировать поиск кластеров: система непересекающихся множеств (DSU), IndexListArray и основанная на них структура Clusters.

Введение

Выделение кластеров связанных частиц, имеющих некоторые характеристики, в молекулярной системе имеет значение для анализа фазовых переходов или стеклования [1,2], идентификации системы водородных связей [3] и других похожих задач молекулярного моделирования. В данной работе рассматривается задача выделения кластеров и создания эффективной структуры данных для хранения информации о принадлежащих к ним частицам. Реализация выполнена на языке Julia как компонент библиотеки анализа траекторий молекулярного моделирования MDPProcessing.jl [4,5].

Алгоритм и организация данных

Под кластером будем понимать компоненту связности графа, вершины которого соответствуют частицам в молекулярной системе, а в качестве ребер рассматривается наличие связи между частицами. Связь в простейшем случае определяется как нахождение частиц на расстоянии меньше порогового друг от друга [2]. С позиции анализа траекторий молекулярного моделирования требуется реализация следующих операций для подобного графа:

1. определение номера (идентификатора) кластера, соответствующего каждой из частиц

2. определение принадлежности двух частиц одному и тому же кластеру

3. получение по идентификатору кластера списка частиц, входящих в него (например, для дальнейшего анализа геометрии)

Первые две операции соответствуют классической структуре данных «система непересекающихся множеств» (англ. disjoint set union, DSU). Эффективные алгоритмы построения соответствующей структуры данных при наличии списка ребер графа приводятся в классических учебниках по компьютерным наукам [6]. Однако для эффективного выполнения операции (3) классическая структура данных должна быть дополнена.

Входные данные представлены в текстовом виде в формате dump из пакета моделирования LAMMPS [7]. Входной файл содержит информацию об анализируемой системе (скорости, положения, идентификаторы частиц и т.п.), геометрии области моделирования (box). При выделении кластеров по критерию близости частиц предполагаются периодические граничные условия: расчётный параллелепипед считается окружённым со всех сторон точно такими же параллелепипедами с идентичными положениями частиц.

Выделение кластеров на основе критерия близости организовано следующим образом. Список пар частиц в системе system, находящихся на расстоянии менее rcut друг от друга, генерируется на лету функцией поиска соседей neighbor_pairs(system, rcut). На его основе создается структура DSU, привязывающая номер частицы к номеру кластера, к которому она относится (cluster ID). В качестве номера кластера выбирается наименьший из номеров входящих в кластер частиц, что соответствует соглашению, принятому при выделении кластеров программой LAMMPS.

Реализация DSU соответствует алгоритму QuickUnion с эвристикой «сжатия путей» (Quick Union with Path Compression, QUPC) [6]. QUPC представляет данные в виде дерева, каждому элементу которого сопоставляется идентификатор родительского элемента. При использовании сжатия путей алгоритмическая сложность добавления новой связи равна $O(\log_2 N)$, где N — число частиц. Таким образом, структура имеет довольно высокую теоретическую эффективность.

В рамках MDPProcessing.jl структура данных DisjointSets используется для хранения системы непересекающихся множеств, генерируемой с помощью QUPC. Для эффективной реализации всех необходимых операций (1-3) после построения структуры DisjointSets создается структура данных Clusters, содержащая поля

```
id::Vector{Int}
list::IndexListArray{1,Int}
state::MDState
```

Поле id соответствует идентификатору кластера для каждой частицы, поле list служит для быстрого получения списка частиц в каждом кластере. Этот список представлен в виде структуры «список ячеек», широко применяемую в алгоритмах поиска со-

седей [8]. В нашем случае, в качестве ячеек рассматриваются выделенные в системе кластеры. Для хранения информации о частицах в кластере применяется $IndexListArray\{N, T < Integer\}$ – структура данных для хранения N -мерного массива связанных списков с индексацией, начинающейся с 0.

Расположение элементов структуры отображено в таблице 1:

Таблица 1. Структура IndexListArray

| Диапазон индексов | 1 | 2:N+1 | N+2:N+ncells+1 | N+ncells+2 | N+ncells+3:end |
|-------------------|-------------|-----------------|---------------------------|----------------------------|------------------|
| Данные | Число ячеек | Размеры массива | Массив головных элементов | Число размещенных объектов | Индексы объектов |

Преимущество использования такой организации данных заключается в возможности получить список частиц, принадлежащих каждому кластеру, не перебирая полностью весь массив идентификаторов кластера id .

Производительность на тестовых задачах

Корректность выделения кластеров проверена путем анализа молекулярной системы с радиусом для связывания частиц, немного превышающим равновесную длину связи в молекуле, но меньше межмолекулярных расстояний. Для примера взят файл dump с моделированием додекана (рис. 1). В этом случае ожидается, что каждый кластер будет соответствовать одной молекуле, что подтвердилось в тесте.

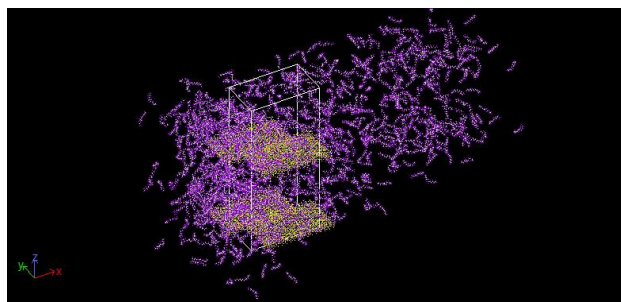


Рис. 1. Трехмерная модель додекана

Для проверки соответствия производительности теоретической оценке времени добавления связи $O(\log_2 N)$ было рассчитано время определения компонент связности графа, заданного синтетическими условиями: количество вершин N задавалось в виде $N = 1000p^3$, а пары соединяемых вершин имели вид $(k, k-1)$, где величины p взяты от 1 до 8, а k пробегает диапазон от 2 до N в порядке возрастания. Такому порядку ввода информации о ребрах графа отвечает алгоритмическая сложность в худшем случае.

Время, затраченное на выделение кластеров, масштабируется с размерами системы примерно линейно, что говорит о корректной реализации алгоритма QUPC (рис.2).

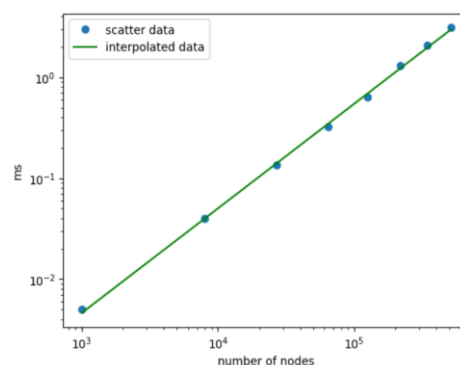


Рис.2. Зависимость времени работы алгоритма QUPC от количества узлов, объединяемых в одну компоненту связности

Заключение

Представленное дополнение к библиотеке анализа траекторий молекулярного моделирования MDProcessing.jl позволяет проводить кластеризацию частиц в молекулярном моделировании по критерию близости частиц. Структура данных для хранения информации о кластерах соответствует классической структуре Disjoint Set Union, дополненной списком частиц в каждом кластере. Итоговый результат работы программы – файл формата dump с выделенными кластерами, которые можно визуализировать в программном обеспечении (например, OVITO).

Время работы алгоритма масштабируется с числом частиц N как $O(N \log_2 N)$, что соответствует теоретической оценке. Корректное выполнение программы проверено на тестах. В работе было проверено, что программная реализация соответствует ожидаемой теоретической эффективности, известной из литературы других авторов, и в ней не содержатся скрытые алгоритмические ошибки.

Список литературы

1. Kirova E. M., Pisarev V. V. Morphological aspect of crystal nucleation in wall-confined supercooled metallic film //J. Phys.: Condens. Matter. – 2021. – V. 33. – P. 034003.
2. ten Wolde P. R., Frenkel D. Numerical study of gas-liquid nucleation in partially miscible binary mixtures //J. Chem. Phys. – 1998. – V. 109. – P. 9919-9927.
3. Родникова М. Н. и др. Об упругости пространственной сетки водородных связей в жидкостях //Журнал физической химии. – 2003. – Т. 77. – №. 10. – С. 1917.
4. MDProcessing.jl library. URL: <https://gitlab.com/pisarevvv/mdprocessing.jl>
5. Pisarev V., Panov M. MDProcessing.jl: Julia Programming Language Application for Molecular Dynamics Trajectory Processing //in: Lecture Notes in Computer Science. – 2023. – V. 14389. – P. 209-222.
6. Sedgewick R., Wayne K. Algorithms. 4th ed. – Addison-Wesley Professional, 2011. 976 P.
7. Thompson A. P. et al. LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales //Comp. Phys. Comm. – 2022. – V. 271. – P. 108171.
8. Allen M. P., Tildesley D. J. Computer simulation of liquids. 2nd ed. – Oxford university press, 2017.

МОДЕЛИРОВАНИЕ РАДИАЦИОННОГО ЗАРЯЖЕНИЯ ПОЛИМЕРНОЙ ИЗОЛЯЦИИ ПРОВОДОВ КОСМИЧЕСКОГО ПРИМЕНЕНИЯ ПРИ НИЗКИХ ТЕМПЕРАТУРАХ

Толстиков С.Ю., Муллахметов И.Р.
 Национальный исследовательский университет
 «Высшая школа экономики»,
 аспирантская школа по техническим наукам

Аннотация

В работе рассмотрена модель возникновения электростатических разрядов в полимерном диэлектрическом слое изоляции проводов космического применения при низких температурах. В качестве исходных параметров были использованы данные по радиационной электропроводности, полученные экспериментально, для таких материалов, как полиимид (ПИ) торговой марки «Картон» и полиэтилентерефталат (ПЭТФ) торговой марки «Mylar» при комнатной и низкой температурах.

Введение

Полимерные материалы находят широкое применение в космической индустрии, особенно в составе элементов космических аппаратов (КА), таких как слои экранно-вакуумной теплоизоляции, изоляция проводов и кабелей, печатные платы и корпуса микроэлектронной аппаратуры. В космическом пространстве эти материалы подвергаются воздействию ионизирующего излучения, в том числе ускоренных электронов, что приводит к дифференциальному зарядению диэлектриков [1-4]. Это, в свою очередь, может вызывать электростатические разряды (ЭСР), способные приводить к необратимым отказам радиоэлектронной аппаратуры космических аппаратов.

С увеличением применения полимеров и ростом числа связанных с этим отказов радиоэлектронной аппаратуры, исследование и предотвращение ЭСР становится особенно актуальным [5].

Основной задачей данного исследования является моделирование процесса возникновения ЭСР в полимерных диэлектрических слоях изоляции проводов КА при низких температурах и под воздействием ионизирующего излучения (условиях приближенном к условиям околоземной плазмы) для оценки физической возможности протекания ЭСР типа «полимерная изоляция - жила».

Новизна исследования заключается в использовании экспериментально полученных данных о радиационной электропроводности ПИ и ПЭТФ при низкой температуре для моделирования зарядения полимерного диэлектрического слоя изоляции проводов КА.

Модель электризации полимерной изоляции проводов космического применения

Для моделирования радиационного зарядения цилиндрического слоя изоляции проводов космического применения (рис. 1) предлагается модель на основе аксиально симметричной математической

модели [6] с некоторыми изменениями. Изначально имеем:

$$F = -\frac{S_0(R_2^2 - R_1^2)}{2\gamma R_1}$$

где F – напряженность электрического поля в диэлектрическом слое провода, $\text{В}\cdot\text{м}^{-1}$; t – время облучения, с; $S_0 = -6\cdot 10^{-5} \text{ А}\cdot\text{м}^{-3}$ – скорость инжекции электронов в объем диэлектрического слоя провода; R_1 – радиус токопроводящего цилиндра (металлическая жила провода), м; R_2 – внешний радиус диэлектрической изоляции провода, м.

В том числе нами впервые были получены экспериментально значения радиационной электропроводности как функции от времени $\gamma_r(t)$ для ПЭТФ и ПИ в условиях протекания магнитосферной суббури ($i_0 = 10^{-5} \text{ А}\cdot\text{м}^{-2}$), как «наихудшего случая» и рассчитаны согласно методике, изложенной в работе [7].

С учетом того, что полная электропроводность материала состоит из темновой и радиационной электропроводностей:

$$\gamma = \gamma_d + \gamma_r$$

А аналитическое выражение полученной в работе [6], не учитывало радиационную составляющую электропроводности. так для моделирования процесса зарядения полимерного диэлектрического слоя изоляции провода с учетом радиационной электропроводности материалов, изменяющейся во времени, выражение было приведено к дифференциальной форме:

$$\frac{dF}{dt} = -\frac{S_0(R_2^2 - R_1^2)}{2(\gamma_d + \gamma_r(t))R_1}$$

где γ_d – величина темновой проводимости полимера, $\text{ом}^{-1}\cdot\text{м}^{-1}$; $\gamma_r(t)$ – радиационная проводимость полимера (в её облучаемой части), $\text{ом}^{-1}\cdot\text{м}^{-1}$.

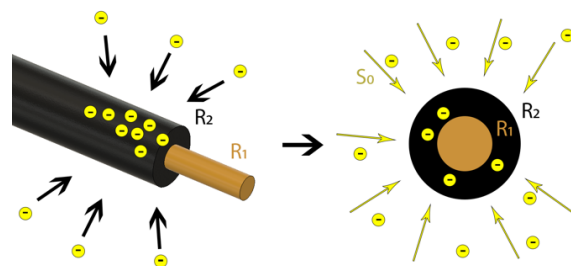


Рис.1. Модель накопления зарядов в диэлектрическом слое провода космического применения

Необходимо подчеркнуть, что в предлагаемой модели значение электрического поля F при радиусе R_2 , т.е. на дальнем цилиндрическом экране, равно нулю. Максимальное значение F (по модулю) достигается при радиусе R_1 .

Значение скорости инжекции электронов в объем диэлектрического слоя провода S_0 , равное $-6\cdot 10^{-5} \text{ А}\cdot\text{м}^{-3}$, было выбрано для «худшего случая» с точки зрения электризации КА в соответствии

с рекомендациями Фредериксона [8], полученными в результате анализа данных с технологического спутника CRRES.

Электризация полимерной изоляции проводов космического применения

Результаты моделирования, проведенные нами согласно нашей модели электризации полимерной изоляции, учитывающей радиационную электропроводность материала этой самой изоляции при низких температурах, представленные на рисунках 2 и 3 показывают, что при низких температурах напряженность поля в полимерных диэлектриках существенно увеличивается, и в отличие от комнатной температуры могут возникать электростатические разряды. Основные параметры исследуемых образцов проводов КА представлены в таблице 1.

Таблица 1. Параметры образцов проводов космического применения

| Провод | Коаксиальный | MC26-15 |
|---|-----------------------|-----------------------|
| Материал изоляции | ПЭТФ | ПИ |
| $\gamma_{db}, \text{ом}^{-1} \cdot \text{м}^{-1}$ | $10^{-15} - 10^{-16}$ | $10^{-12} - 10^{-13}$ |
| $R_1, \text{м}$ | 0,01 | 0,0004 |
| $R_2, \text{м}$ | 0,0254 | 0,000575 |

Результаты моделирования показывают, что эффект, наблюдаемый на пленках при низких температурах [9], проявляется в полной мере и при использовании материалов ПЭТФ и ПИ в качестве изоляции проводов КА.

Результаты натурных испытаний на спутнике SCATHA (NASA) показали, что ЭСР возникают при достижении значения электрического поля в объеме диэлектрика $2 \cdot 10^7 \text{ В} \cdot \text{м}^{-1}$. Эта величина служит как критерием, так и порогом [5], после которого возможно возникновение электростатических разрядов.

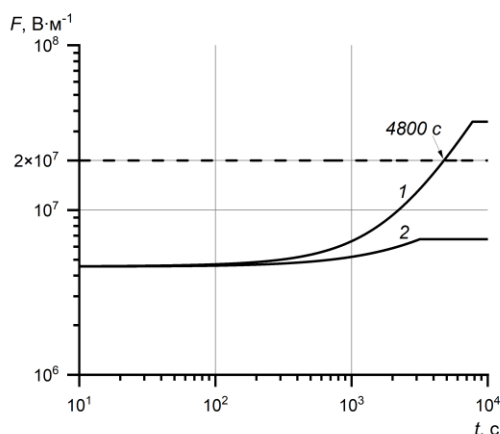


Рис.2. Расчетные значения напряженности в диэлектрическом слое (ПЭТФ) коаксиального провода от времени облучения при 103 К (1) и 298 К (2)

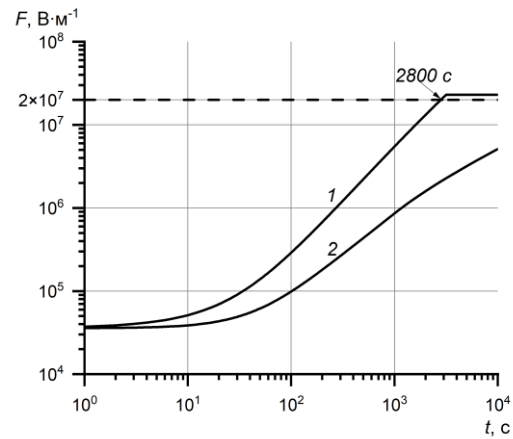


Рис.3. Расчетные значения напряженности в диэлектрическом слое (ПИ) провода MC26-15 от времени облучения при 103 К (1) и 298 К (2)

Таким образом, из графиков видно, что возможность возникновения ЭСР при низких температурах в условиях суббури присутствует у обоих материалов. Спустя 4800 с для ПЭТФ и 2800 с для ПИ при достижении напряженности поля выше заданной критериальной величины $2 \cdot 10^7 \text{ В} \cdot \text{м}^{-1}$, появляются физические условия, при которых могут возникать ЭСР типа «изоляция» – «токопроводящая жила».

Заключение

Проведена оценка физической возможности возникновения ЭСР в полимерном диэлектрическом слое изоляции проводов в условиях, приближенных к условиям горячей околоземной плазмы. Показано, что выбор материала для изоляции проводов должен учитывать как условия внешней радиационной обстановки, так и специфику поведения материала при низких температурах. Это подчеркивает важность моделирования поведения материала изоляции еще на стадии подготовки полётных миссий и проектирования КА, особенно с учетом современных требований к надежности в космической технике.

Так, характерное время магнитосферной суббури составляет приблизительно час, ввиду этого на основании полученных результатов рекомендуется применение ПЭТФ в качестве изоляционного материала для проводов КА, что способствует понижению вероятности возникновения ЭСР, в отличие от ПИ.

Список литературы

1. Frederickson A.R. Electric Discharge Pulses in Irradiated Solid Dielectrics in Space // IEEE Transactions on Electrical Insulation. – 1983. – vol. 18. – N. 3. – P. 337–349.
2. Акишин А.И. Электроразрядные сбои в космических аппаратах в зоне космических излучений. // Перспективные материалы. – 2010. – № 2. – С. 27–32.
3. Catani J.-P., Payan D. Electrostatic behavior of materials in a charging space environment / Proc. 9th Int. Symp. On Materials in a Space Environment. – Noordwijk: ESA Publ. Division, 2003. – P. 3-16.

4. Тютнев А.П., Ванников А.В., Мингалеев Г.С. Радиационная электрофизика органических диэлектриков. – М.: Энергоатомиздат, 1989. – 192 с.

5. NASA-Technical Handbook: Mitigating in-Space Charging Effects – A Guideline: document Rec. NASA-HDBK-4002A – USA: NASA, 2017. – 180 p.

6. Толстикова С.Ю., Саенко В.С., Тютнев А.П. Исследование и разработка безразрядной изоляции проводов космического применения // Журнал технической физики. – 2022. – Т. 92. – № 6. – С. 779–782.

7. Tyutnev A.P., Saenko V.S., Mullakhmetov I.R., Pozhidaev E. D. Experimental and theoretical investigations of the radiation-induced conductivity in spacecraft polymers at extremely low temperatures // Journal of Applied Physics. – 2023. – vol. 134. – Article 095903.

8. Frederickson A.R., Brautigam D.H. Mining CRRES IDM Pulse Data and CRRES Environmental Data to Improve Spacecraft Charging/Discharging Models and Guidelines. Unpublished report. – NASA contract NAS-1407, task order 10676, final and fourth quarterly report. – September 30, 2003.

9. Муллахметов И.Р., Саенко В.С., Тютнев А.П., Пожидаев Е.Д. Низкотемпературная радиационная электропроводность полистирола под действием электронов низких энергий // Журнал технической физики. 2023. Т. 93. № 1. С. 130-134.

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПОИСКА КОЛЛИЗИЙ В ГЕОМЕТРИЧЕСКИХ МОДЕЛЯХ ПРИ РАЗНЫХ СПОСОБАХ ПРЕДСТАВЛЕНИЯ ОБЪЕКТОВ

Ершов И.С.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ*

Аннотация

Геометрические модели широко используются при проектировании сооружений. При работе с моделью возможно появление пересечений, называемых коллизиями, наличие которых влечет за собой негативные последствия, приводящие к дополнительным издержкам на их устранение. Вследствие чего возникает потребность в регулярных проверках геометрической модели на коллизии. В данной работе методами вычислительного эксперимента проведен анализ скорости построения пересечений между объектами модели, заданными с помощью различных способов представления геометрии. Сформулированы рекомендации по выбору способа представления в зависимости от особенностей используемых геометрических объектов.

Введение

При проектировании архитектурных сооружений используются информационные модели зданий. Для того, чтобы создать информационную модель, при-

меняются различные технологии, одна из которых называется BIM (Building information model - информационная модель здания) [1]. Модели содержат информацию о геометрических, физических и функциональных характеристиках сооружений. При проектировании информационной модели могут возникать коллизии, которые приводят к значительным затратам, при их обнаружении на этапе подготовки итоговых чертежей или демонтажа. Вследствие чего при проектировании сооружения происходит регулярная проверка модели на наличие коллизий.

Цифровая модель зданий представляет из себя совокупность геометрических объектов, которые могут быть представлены разными способами. Наиболее распространенными являются треугольные сетки [2] и B-гер [3]. Треугольная сетка – это представление поверхности трехмерного объекта с помощью набора треугольников, в свою очередь B-гер – это представление, использующее в качестве примитивов точки, кривые и поверхности. Целью данной работы является анализ эффективности поиска коллизий для треугольных сеток и B-гер с использованием библиотеки для вычислительной геометрии, называемой CGAL [4], в которой содержатся структуры данных и алгоритмы необходимые для вычисления перечислений для приведенных представлений.

Треугольная сетка

Одним из способов представления геометрии объектов является использование треугольных сеток. С помощью набора треугольников можно аппроксимировать различные трехмерные поверхности с необходимой точностью.

Одним из способов организации данных о представлении в виде сетке является структура half-edge, которая хранит информацию о гранях, ребрах и точках [5]. Ребро разбивается на два полуребра, проассоциированных друг с другом, относящихся к разным граням и имеющих разные направления. На рис.1 представлена схема структуры half-edge, где F - грань, v - вершина и e - полуребро, соответствующей грани, в которой лежит, и вершине, из которой выходит. Обход такой структуры данных происходит по граням, используя информацию о соседних гранях, хранимую в ребрах.

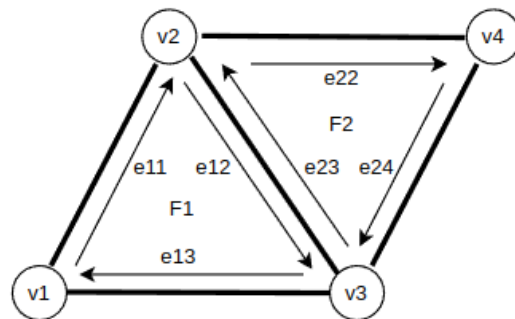


Рис.1. Представление half-edge

В библиотеке CGAL такой способ организации данных реализован в классе Surface_mesh. Особен-

ность реализации заключается в том, что вместо указателей на рёбра используется индексация.

В документации библиотеки CGAL описан алгоритм, позволяющий находить пересечение между двумя сетками. Для работы алгоритма необходимо, чтобы передаваемые на вход алгоритма сетки обладали следующими свойствами: отсутствие самопересечений и наличие замкнутости у сетки. Отсутствие самопересечения заключается в отсутствии пересечений между гранями внутри одной сетки. В свою очередь, сетка является замкнутой, если ограничивает собой некоторый объём.

Алгоритм построения пересечений состоит из двух этапов, которые изображены на рис.2. На первом этапе между парой сеток находятся пересекающиеся грани и на их месте добавляются новые вершины, ребра и грани. В результате они образуют замкнутый контур, принадлежащий обеим сеткам. На следующем этапе формируется объект пересечения: он состоит из контура и сеток, рассматриваемых объектов по обе стороны от контура.

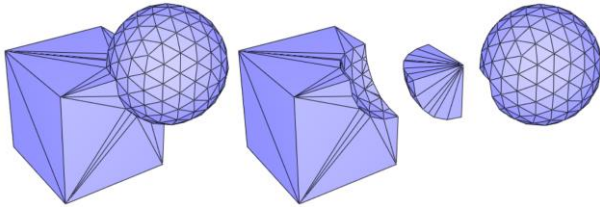


Рис.2. Вычисление пересечения

Таким образом, построение объектов коллизии возможно, если объекты, для которых происходит расчёт коллизии, представимы в виде не самопересекающихся замкнутых сеток.

В-rep

В качестве альтернативного способа представления трехмерных объектов применяется твердотельное моделирование. При таком представлении геометрические объекты задаются с помощью логических операций над конечным набором примитивов [6]. В данной работе рассматривается представление В-гер или граничное представление, которое в качестве примитивов использует кривые, поверхности и точки.

В библиотеке CGAL такой способ организации данных реализован в классе Nef_polyhedron [7]. Данная структура реализована с использованием локальных окрестностей вершин, представляющих из себя проекции вершин на поверхность трехмерной сферы. С помощью данных проекций определяется часть пространства, которую вершина ограничивает. В совокупности пересечение ограниченных пространств образует объект. На рис.3 представлена схема структуры, где v - вершины с окрестностями. Реализация В-гер с использованием Nef_polyhedron позволяет гарантировать замкнутость пересечений при работе с многогранниками.

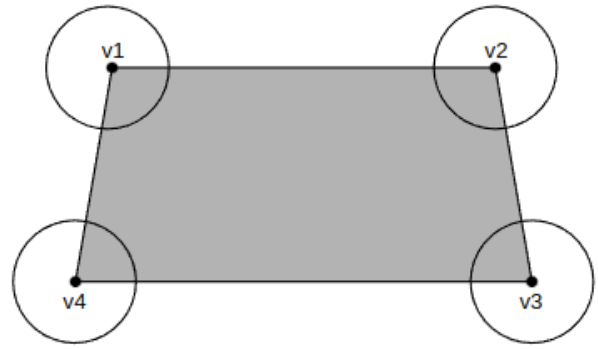


Рис.3. Структура данных В-rep

В качестве входных данных для построения объекта представления Nef_polyhedron применяется замкнутая полигональная сетка, состоящая из набора вершин и граней, где гранью может быть произвольный двумерный многоугольник [8].

При этом существуют ограничения на входные данные при создании Nef_polyhedron. Структура содержит функционал для построения объекта путем пересечения плоскостей, однако ограниченность объекта в таком случае не гарантируется. Кроме того, для структуры применяются точные геометрические вычисления над предикатами для построения объекта Nef_polyhedron. Это приводит к дополнительной обработке входных данных.

Таким образом, с помощью В-гер можно задать представление геометрической модели и вычислять пересечения, если соблюдаются ограничения, накладываемые при построении пересечений с помощью Nef_polyhedron.

Анализ эффективности поиска коллизий

На вход вычислительного эксперимента были поданы пары правильных призм с радиусом основания $R = 5$ и высотой $H = 20$. Количество боковых граней призмы задавалось с помощью параметра N . Выбранный способ параметризации призмы отображен на рис.4.

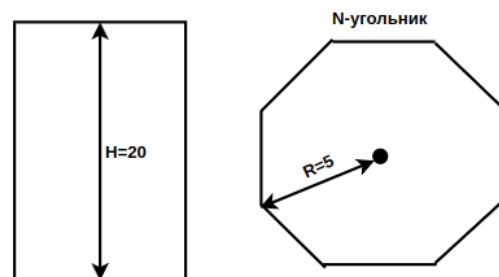


Рис.4. Параметры призмы

Рассмотрим ситуацию, где одна призма смещена в плоскости основания на 2.5 и вдоль оси на 5, как показано на рис.5. При таком расположении призм происходит пересечение поверхностей оснований и боковых граней.



Рис.5. Пересекающиеся призмы со смещением

Количество боковых граней призмы варьируется от 5 до 250 с шагом 10. В результате получена зависимость времени вычисления коллизии от количества граней, проиллюстрированная на рис.6.

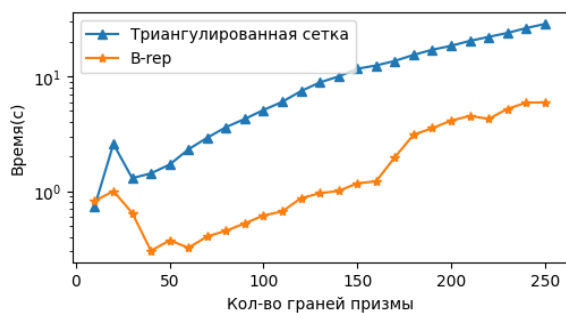


Рис.6. Сравнение времени расчета коллизий для призм, расположенных со смещением

Скорость вычислений для представления геометрической модели с помощью V-гер значительно выше, чем для треугольной сетки. Для призм с количеством граней равным 250 вычисление пересечения происходит почти в 5 раз медленнее V-гер. Это связано с тем, что количество треугольников в основании призмы, заданное с помощью треугольной сетки, прямо пропорционально количеству граней.

Аналогичный результат продемонстрирован при пересечении призм, расположенных путем смещений в плоскости основания на 5, вдоль оси на 10 и поворота вокруг оси x на 30 градусов, как показано на рис.7.

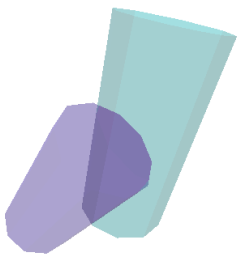


Рис.7. Пересекающиеся призмы под углом

На графике, приведенном на рис.8, видно, что результат измерений аналогичен полученному ранее для смещённых призм.

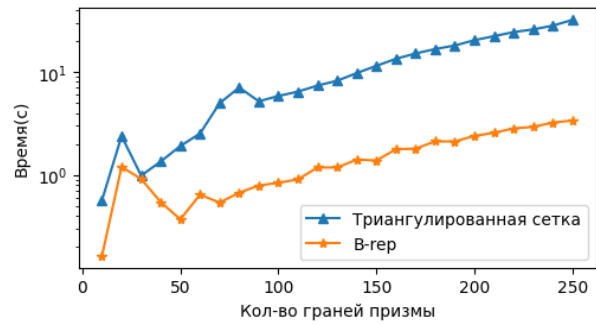


Рис.8. Сравнение времени расчета коллизий для призм, расположенных под наклоном

Рассмотрим ситуацию, где две призмы располагаются перпендикулярно друг относительно друга, как показано на рис.9. При таком расположении в пересечении лежат только боковые грани.



Рис.9. Пересекающиеся призмы

В таком случае количество боковых граней призмы варьировалось от 100 до 500 с шагом 10. Полученная зависимость времени от количества граней отображена на рис.10.

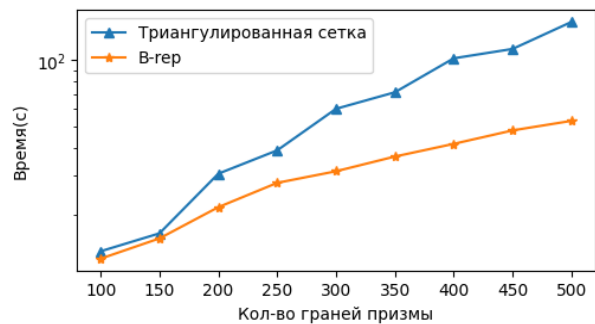


Рис.10. Сравнение времени расчёта коллизий для призм, расположенных перпендикулярно

Эффективность использования V-гер при работе с элементами, количество граней которых меньше 150 сравнимо с эффективностью использования треугольных сеток. Однако при большем количестве граней эффективность V-гер снова возрастает, как это было отмечено в описанных ранее вычислительных экспериментах.

Заключение

В результате анализа двух подходов представления геометрических моделей было выявлено, что для достижения наибольшей производительности поиска коллизий в геометрических моделях необходимо использовать V-гер. Эффективность поиска коллизий

между треугольными сетками демонстрирует схожие результаты только для моделей, в которых использование многогранных стержневых элементов сведено к минимум. При выполнении данного условия треугольные сетки будут предпочтительней V-ger в силу меньшего количества ограничений, накладываемых на входные данные.

Список литературы

1. Червова, Н.А. Коллизии инженерных систем при проектировании в BIM платформах / Н.А. Червова, Д.О. Лепешкина // Строительство уникальных зданий и сооружений. – 2018. – № 3. – С. 19-29. – ISSN 2304-6295
2. Geometric modeling based on triangle meshes / Mario Botsch, Mark Pauly, Christian Rössl [и др.] // SIGGRAPH 2006 - ACM SIGGRAPH 2006 Courses. – 2006. – С. 1-es.
3. Parvin, B B-rep Object Description from Multiple Range Views* / B Parvin, G Medioni // International Journal of Computer Vision. – 1996. – № 20. – С. 81-112.
4. The Computational Geometry Algorithms Library : сайт. – URL: <https://www.cgal.org/> (дата обращения: 13.02.2024)
5. McGuire M. The half-edge data structure //Website: http://www.flipcode.com/articles/article_halfedgepf.shtml. – 2000.
6. Shapiro V. Solid Modeling //Handbook of computer aided geometric design. – 2002. – Т. 20. – С. 473-518.
7. Bieri H. Nef polyhedra: A brief introduction. – Springer Vienna, 1995. – С. 43-60.
8. Botsch M. et al. Polygon mesh processing. – CRC press, 2010.

ПРЕДСКАЗАНИЕ ВЯЗКОСТИ УГЛЕВОДОРОДОВ С ПОМОЩЬЮ МОЛЕКУЛЯРНОЙ ДИНАМИКИ И МАШИННОГО ОБУЧЕНИЯ

Мкртчян Я.Р., Смирнов Г.С.

*Национальный исследовательский университет
«Высшая школа экономики»,
международная лаборатория
суперкомпьютерного атомистического
моделирования и многомасштабного анализа*

Аннотация

В данной работе предлагается расширить метод вычисления вязкости органических соединений, основанный на проведении молекулярно-динамических расчетов и построении машинной модели.

Метод заключается в построении QSPR модели (модели структура-свойство) с использованием регрессии гауссова процесса (GPR) и молекулярных дескрипторов в качестве предикторов для нашей модели. Молекулярные дескрипторы считаются из равновесных траекторий молекул, полученных в результате проведения молекулярно-динамических расчетов.

Введение

Смазочные масла играют важную роль в работоспособности различных механических систем, таких как: двигатели внутреннего сгорания, тормозные системы и так далее. В основном вещества, используемые для таких целей, состоят из базовых масел различной вязкости. В качестве базовых масел используют сложные углеводороды и их смеси. Вязкость – одно из важнейших свойств масла, которое напрямую влияет на трение, стойкость к износу и общее время жизни механизмов.

Так как экспериментальные данные по некоторым веществам не всегда доступны, а проведение эксперимента самостоятельно - трудоёмкая задача, то использование машинной модели поможет избежать этих проблем.

Большинство реализованных моделей машинного обучения используются для предсказания температурной зависимости вязкости. В промышленных целях необходимо знать еще зависимость от давления, поэтому в представленной работе предлагается машинная модель, предсказывающая вязкость при различных температурах и давлениях.

Разработана гауссова-регрессионная модель, способная предсказывать вязкость органических соединений не только при высоких температурах, но и при высоких давлениях. В качестве предикторов используются статические и динамические молекулярные дескрипторы [1].

Метод молекулярной динамики

Метод молекулярной динамики – компьютерный эксперимент, который основан на численном интегрировании системы уравнений движений частиц исследуемой системы. Результатом такого расчета являются траектории частиц, полностью определяющие данную систему [2].

Динамика взаимодействующих частиц задается системой уравнений:

$$m_i \frac{d^2 r_i}{dt^2} = - \frac{\partial U(r_1, \dots, r_N)}{\partial r_i},$$

$$F_i(r_1, \dots, r_N) = - \frac{\partial U(r_1, \dots, r_N)}{\partial r_i},$$

где m_i – масса i -ой частицы, r_i – радиус-вектор i -ой частицы, а функция $U(r_1, \dots, r_N)$ – потенциал, описывающий взаимодействие в системе.

Наиболее распространенная схема численного интегрирования уравнений движения – скоростная схема Верле:

$$v_i \left(t + \frac{\Delta t}{2} \right) = v_i(t) + a_i(t) \frac{\Delta t}{2},$$

$$r_i(t + \Delta t) = r_i(t) + v_i \left(t + \frac{\Delta t}{2} \right) \Delta t,$$

$$v_i(t + \Delta t) = v_i \left(t + \frac{\Delta t}{2} \right) + a_i \left(t + \Delta t \right) \frac{\Delta t}{2},$$

где $a_i(t) = \frac{F_i(t)}{m_i}$ - ускорение i -ой частицы.

Молекулярные дескрипторы

Проведено более 250 молекулярно-динамических расчетов для различных органических соединений в широком диапазоне температур и давлений. На основе полученных траекторий вычислялись статические и динамические молекулярные дескрипторы.

Молекулярные дескрипторы – числовые параметры, характеризующие термодинамические и химические свойства молекул и состояние, в котором находится вещество (состояние характеризуется давлением p и температурой T) [3].

В качестве предикторов были выбраны геометрические дескрипторы, характеризующие пространственное расположение атомов в молекуле и динамические дескрипторы, вычисляющиеся непосредственно из равновесных траекторий движения молекул (являются характеристикой силового поля) [4].

Машинная модель

Одним из подходов для нахождения неизвестной функции f , которая будет способна ставить в соответствие входному вектору из дескрипторов значение переменной отклика (вязкости), являются модели машинного обучения, основанные на регрессии гауссовых процессов (GPR).

$$y = f(\mathbf{x}) + \varepsilon,$$

где \mathbf{x} – входной вектор из дескрипторов, y – переменная отклика (вязкость).

Гауссовы процессы – это обобщение вероятностного распределения Гаусса, в котором вероятностное распределение отвечает за случайные величины, которые являются скалярами или векторами, а стохастический процесс отвечает за характеристики искомой функции f .

Регрессия гауссова процесса – алгоритм машинного обучения, который в качестве предикторов использует небольшой набор параметров. Так как GPR является непараметрической, это позволяет эффективно решать задачи обучения с учителем даже с небольшим количеством данных.

Наиболее важной частью в оптимизации модели был выбор ковариационной функции. Ковариационная функция взвешивает вклады каждой из точек обучающей выборки в прогнозируемую величину, то есть полностью определяет поведение процесса. В процессе оптимизации ядра (ковариационной функции) было выявлено, что оптимальное ядро – Matern:

$$k(x_i, x_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{l} d(x_i, x_j) \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}}{l} d(x_i, x_j) \right),$$

где $d(\cdot, \cdot)$ – евклидово расстояние, $K_\nu(\cdot)$ – модифицированная функция Бесселя и $\Gamma(\cdot)$ – это гамма-функция Эйлера. Оптимальное ядро выбиралось из условия минимума среднеквадратичной ошибки (MSE). Выбиралось ядро с лучшей предсказательной способностью.

Работоспособность модели была проверена на экспериментальных данных по различным углеводородам [5].

Результаты

С помощью F-test, корреляционной матрицы, регрессии Lasso и физических соображений, основанных на анализе теоретических и эмпирических формул вычисления вязкости были выбраны наиболее важные молекулярные дескрипторы, к которым можно отнести плотность, парную энтропию, потенциальную энергию системы.

Обученная модель тестировалась на углеводородах ($C_5 - C_{32}$) и показала хорошую предсказательную способность несмотря на различия в строении соединений и широкий диапазон температур (270 – 500 K) и давлений (1 – 2500 атм), представленный в тестовой выборке.

Заключение

Машинное обучение с использованием статических дескрипторов в качестве предикторов не позволяет построить корректную зависимость вязкости углеводородов от давления. Высокой предсказательной силы модели удастся добиться только при использовании динамических дескрипторов, вычисленных с использованием метода молекулярной динамики.

Разработана машинная модель на языке программирования Python, способная определять вязкость жидкостей при различных условиях, основываясь на структурных свойствах молекул, рассчитанных с помощью молекулярного моделирования.

Список литературы

1. Panwar P., Yang Q., Martini A. Temperature-dependent density and viscosity prediction for hydrocarbons: machine learning and molecular dynamics simulations //Journal of Chemical Information and Modeling. – 2023.
2. Panwar P., Yang Q., Martini A. PyL3dMD: Python LAMMPS 3D molecular descriptors package //Journal of Cheminformatics. – 2023. – Т. 15. – №. 1. – С. 69.
3. Норман Г. Э., Стегайлов В. В. Стохастическая теория метода классической молекулярной динамики //Математическое моделирование. – 2012. – Т. 24. – №. 6. – С. 3-44.
4. Nikitiuk B. I. et al. Pair entropy and universal viscosity scaling for molecular systems via molecular dynamics simulations //Journal of Molecular Liquids. – 2022. – Т. 368. – С. 120714.
5. Meng X. et al. Viscosity measurements of ortho-xylene, meta-xylene, para-xylene and ethylbenzene //The Journal of Chemical Thermodynamics. – 2016. – Т. 95. – С. 116-123.

МОДЕЛИРОВАНИЕ И АНАЛИЗ КОНСТРУКЦИИ МИКРОФЛЮИДНОГО ЧИПА

Варфоломеева А.А., Паршина С.Р., Пятко Л.А.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ

Аннотация

Работа посвящена построению и анализу модели напряженно-деформированного состояния мембраны микрофлюидного чипа. Мембрана, являясь наиболее подвижным элементом чипа, подвергается механическим деформациям и напряжениям, которые способствуют накоплению микродефектов и повышению риска просачивания воздуха в его каналы. С помощью компьютерной модели исследуется влияние возможных изменений конструкции мембраны на минимизацию напряжений.

Введение

Микрофлюидный чип — это компактное устройство широко используемое в научных и медицинских целях для изучения биологических объектов, основанное на микрофлюидных платформах и культивировании клеток *in vitro* [1-3]. Устройство представляет собой систему микроскопических каналов и полостей, позволяющую реализовывать различные операции выделения, сортировки и анализа биологической пробы. Микроканалы формируются в пластине полидиметилсилоксана (PDMS), которая наклеивается на предметное стекло. Для обеспечения движения жидкости внутри каналов используется система пневмонасосов, мембраны которых также формируются в PDMS пластине.

В процессе эксплуатации воздух из пневмосистемы просачивается через мембраны, что приводит к образованию пузырьков внутри микрофлюидной системы, которые затрудняют прокачку по ней жидкости [4, 5]. Одной из возможных причин может быть износ PDMS, приводящий к появлению микродефектов, пропускающих воздух. Наибольшему износу за счет наличия концентраторов механических напряжений подвергаются участки стыковки мембран к основному массиву PDMS.

В работе представлены компьютерная модель мембраны микрофлюидного чипа, которая позволяет проанализировать напряженно-деформированное состояние мембраны, выявить наличие концентраторов напряжений и предложить конструктивные решения по их устранению.

Принцип работы устройства

Микрофлюидная система — это компактное устройство, в котором малые объемы жидкости перемещаются по каналам размерами в сотни микрон. Чип представляет собой пластину из стекла, кремния или полимерных соединений, в которой отлиты микроканалы (см. рис.1). Каналы имеют разную форму, а также могут соединяться вместе для достижения

разных целей: смешивания и разделения потоков, сортировки клеток.

Жидкость управляется с помощью пневмонасосов, важной составляющей которых являются мембраны (см. рис.2). В верхней камере клапана создается давление, и с помощью движения мембран жидкость начинает циркулировать по каналам с необходимой скоростью.

Конструкция микрофлюидного чипа зависит от конкретной задачи, которую пытаются решить исследователи-биологи. В дальнейшем будет использоваться прототип микрофлюидного устройства, разрабатываемый лабораторией микрофлюидных технологий ИБХ РАН. Контур циркуляции этого чипа приведен на рис.1. Питательная среда внутри контура приводится в движение при помощи насоса-маршрутизатора состоящего из трех рабочих клапанов (см. рис.1.2), обеспечивающих пульсирующее движение жидкости в любом направлении контура. Демпферы (см. рис.1.3) представляют собой широкие камеры, предназначенные для гашения волн жидкости и равномерного протекания питательной среды около ячеек с биологическими пробами (см. рис.1.4)

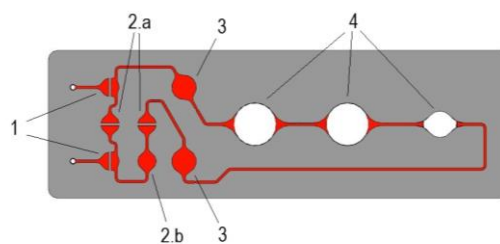


Рис.1. Схема каналов микрофлюидного чипа, состоящая из клапанов подачи/сброса питательной среды (1), рабочих клапанов с перегородкой (2a) и без перегородки (2b), демпфирующих мембран (3) и ячеек с клетками (4)

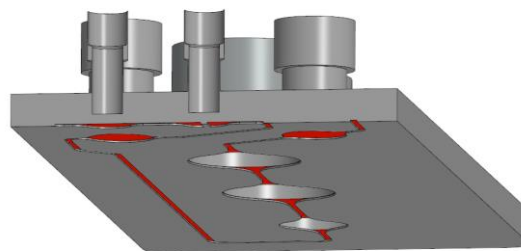


Рис.2. Разрез слоя PDMS микрофлюидного чипа

Процедура построения модели

Механические свойства PDMS сильно зависят от деталей процесса его производства. Обычно используется следующая технология: смешивается эластомер с отвердителем, затем происходит удаление пузырьков в вакуумной камере, после чего смесь заливается в форму с системой каналов, где застывает при высокой температуре. При изменении пропорции смешивания основы PDMS с отвердителем, а

также времени и температуры отвердевания упругие свойства материала сильно меняются [5, 6]. Поскольку в работе изучается реальный прототип микрофлюидного чипа с известными размерами, технологией производства и режимом работы, для дальнейшего анализа будет использован материал, получаемый в результате отверждения в течение 1 часа при температуре 80°C в соотношении 10:1. Упругие свойства PDMS, получаемые при такой технологии производства, определяются следующими характеристиками: коэффициент Пуассона $\sigma = 0,495$ и модуль Юнга $E = 0,6$ МПа [6, 7]. Радиус нижней камеры - 1,5 мм, радиус верхней камеры - 1 мм, толщина мембраны 0,6 мм, высота каналов 0,1 мм. режим работы насоса подразумевает изменение давления в пределах от -30 до 30 кПа относительно атмосферного давления, создавая соответствующее давление в верхней камере. Давление в нижней камере принимается равным атмосферному. Трение между мембраной и стеклом считается нулевым, поскольку не вносит существенного вклада в исследуемые характеристики.

Для анализа напряженно-деформированного состояния мембраны был использован метод конечных элементов, реализованный в программе Abaqus. В рассматриваемой задаче объектом моделирования являлся клапан и прилегающая к нему часть основного слоя PDMS.

Результаты моделирования

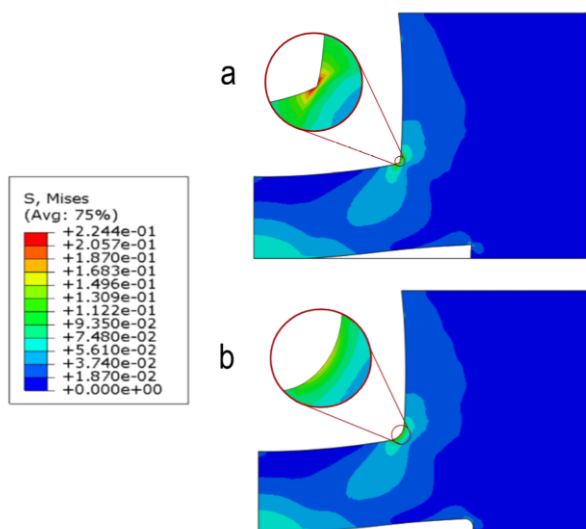


Рис.3. Тепловая карта напряжения по Мизесу а) до скругления, б) при скруглении радиусом 70 мкм

По результатам моделирования работы клапана можно заметить концентраторы напряжения в слое PDMS. Он образуется на углу стыковки мембраны и основного массива PDMS (см. рис.3а).

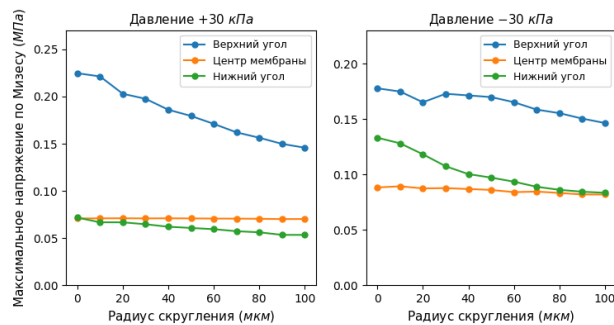


Рис.4. Зависимость максимального напряжения по Мизесу от радиуса скругления стыков

Предложенной модификацией конструкции чипа является сглаживание углов стыковки мембраны для достижения более равномерного распределения напряжения (см. рис.3б). Были рассмотрены различные варианты радиуса скругления от 0 мкм (прямой угол) до 100 мкм. По графикам (см. рис.4) видно, что при увеличении угла скругления, максимальное напряжение убывает в обоих состояниях работы насоса. У клапана с радиусами скругления равными 100 мкм максимальное напряжение по Мизесу уменьшается на 35,2% в состоянии положительного давления, и на 17,6% в состоянии отрицательного давления.

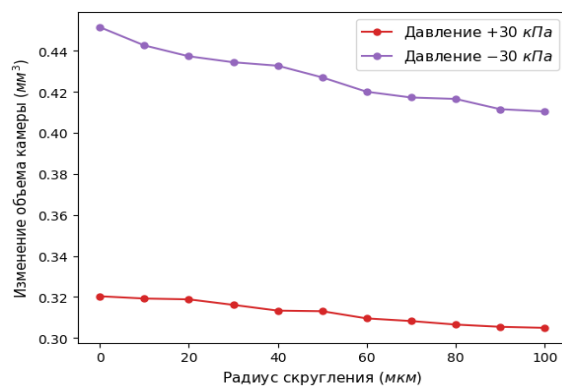


Рис.5. Зависимость измененного объема от радиуса скругления

Одним из недостатков любой модификации конструкции чипа является изменение способности жидкости циркулировать так же, как и до корректировок. Основной величиной, влияющей на перекачивание жидкости внутри каналов, является изменение объема внутри камеры под мембраной во время работы насоса. Как показано на рис.5, при увеличении радиуса скругления, изменение объема уменьшается в обоих граничных состояниях работы.

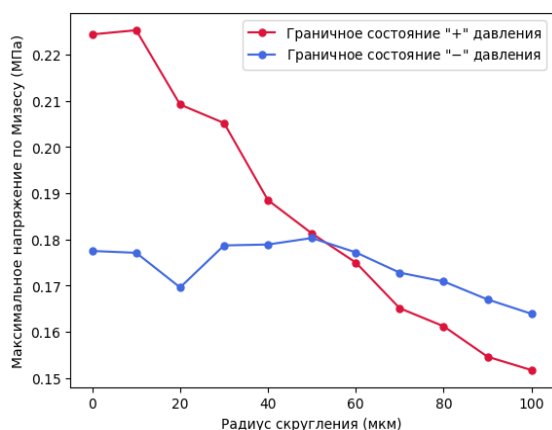


Рис. 6. Максимальное напряжение по Мизесу после изменения давления, в зависимости от радиуса скругления

Для того чтобы измененный объем оставался прежним и скругление угла не влияло на способность насосов перемещать жидкости, можно увеличить границы давления в рамках которых работает насос. После такого изменения давления увеличиваются механические напряжения, но взамен питательная среда будет циркулировать также, как и до скругления углов стыковки. Как видно на рис.6, максимальное напряжение в обоих граничных состояниях работы будет наименьшим при радиусе скругления 100 мкм. Так, для "+" давления напряжение по Мизесу уменьшается с 0,2244 МПа до 0,1517 МПа (на 32,3%), а для "-" давления оно уменьшается с 0,1775 МПа до 0,1639 (на 7,7 %).

Заключение

В работе представлена компьютерная модель мембраны и исследована зависимость максимального напряжения от конструкции чипа. За счет скругления углов на участке стыковки мембраны к основному массиву PDMS, удалось уменьшить концентраторы напряжения. Дальнейшая работа будет направлена на построение компьютерной модели течения жидкости по системе каналов чипа с использованием клапана Теслы и систематизации результатов для рекомендаций по выбору параметров мембраны.

Список литературы

1. Tataru, A.M. Role of Tissue Engineering in COVID-19 and Future Viral Outbreaks / A.M. Tataru // Tissue Eng.. - 2020. - Vol. A № 26. - С. 468–474.
2. Breakthroughs and Applications of Organ-on-a-Chip Technology / Mufeeda C. Koyilot, Priyadarshini Natarajan, Clayton R. Hunt, Sonish Sivarajkumar, Romy Roy, Shreeram Joglekar, Shruti Pandita, Carl W. Tong, Shamsudheen Marakkar, Lakshminarayanan Subramanian, Shalini S. Yadav, Anoop V. Cherian, Tej K. Pandita, Khader Shameer, Kamlesh K. Yadav // Cells. - 2022. - №11. - С. 1828.
3. Mehling, M. Microfluidic cell culture / Matthias Mehling, Savas, Tay // Current Opinion in Biotechnology. - 2014. - №25. - С. 95–102.

4. Bubble inclusion and removal using PDMS membrane-based gas permeation for applications in pumping, valving and mixing in microfluidic devices / Michael Johnson, Greg Liddiard, Mark Eddings, Bruce Gale // Journal of Micromechanics and Microengineering. - 2009. - № 19. - С. 1-9.

5. Lamberti, A. PDMS membranes with tunable gas permeability for microfluidic applications / A. Lamberti, S. L. Marassoab, M. Cocuzzaab // RSC Adv.. - 2014. - № 4. - С. 61415.

6. The relationship between the Young's modulus and dry etching rate of polydimethylsiloxane (PDMS) / Matthew L. Fitzgerald, Sara Tsai, Leon M. Bellan, Rebecca Sappington, Yaqiong Xu, Deyu Li // Volume. - 2019. - №21. - С. 26.

7. Müller, A. A quick and accurate method to determine the Poisson's ratio and the coefficient of thermal expansion of PDMS / Angelina Müller, Matthias C. Waplera, Ulrike Wallrabe // Soft Matter. - 2019. - №15. - С. 779-784.

ВЛИЯНИЕ СВЯЗАННЫХ СОСТОЯНИЙ НА СВЕРХПРОВОДИМОСТЬ В МОДЕЛИ ХАББАРДА НА ЛЕСТНИЧНЫХ РЕШЕТКАХ

Пиле Я.Э.

Национальный исследовательский университет
«Высшая школа экономики»,
лаборатория вычислительной физики
МИЭМ НИУ ВШЭ

Аннотация

В работе показано, что на лестничных решетках малой ширины в модели Хаббарда при спин-зависимой амплитуде туннелирования наличие связанных состояний трех фермионов (тримеров) экспоненциально подавляет сверхпроводящие FFLO-корреляции при соизмеримых плотностях носителей заряда $2n_{\uparrow} = n_{\downarrow}$. Однако уже при ширине лестничной решетки равной 5 тримеры не оказывают на сверхпроводящие корреляции никакого значимого влияния.

Введение

Последние достижения в области исследования ультрахолодных атомных газов открывают новые перспективы для решения фундаментальных теоретических проблем в прямых экспериментах [1]. Одним из все еще открытых вопросов является существование сверхпроводимости при несбалансированном количестве фермионов с различными направлениями спина. Одной из возможностей является знаменитое состояние Фульде-Феррелла-Ларкина-Овчинникова (FFLO) [2,3], в котором сверхпроводящий параметр порядка становится модулированным в пространстве. Экспериментальный поиск сверхпроводимости в поляризованных атомных квантовых газах пока ограничивается трехмерными конфигурациями [4]. Многообещающим направлением является удержание атомов в вытянутых ловушках, где

состояние FFLO, является очень устойчивым [5], что подтверждено детальным численным моделированием, как для строго одномерной, так и для квазиодномерной конфигурации [6,7].

Другой интересной темой, которая в настоящее время исследуется экспериментально, является формирование связанных состояний в ферми-газах с неравными массами, таких как смеси ${}^6\text{Li}$ и ${}^{40}\text{K}$, вблизи резонанса Фешбаха [8,9]. Другой экспериментальной реализацией рассматриваемой модели может быть двухкомпонентный ферми-газ в спин-зависимой оптической решетке, устроенной так, чтобы соответствующие эффективные массы были разными [10].

Модель

В работе рассматривается модель Хаббарда с притягивающим взаимодействием на лестничной решетке ширины W и длины L со спин-зависящими амплитудами туннелирования.

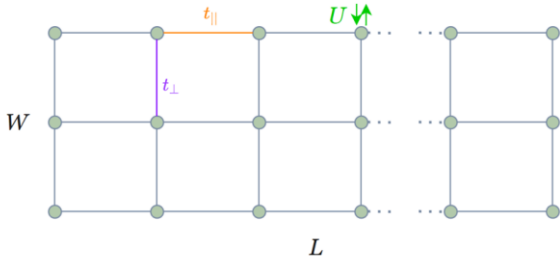


Рис.1. Схематическое изображение лестничной решетки

Гамильтониан имеет вид:

$$\hat{H} = -t_{(\parallel,\perp),\sigma} \sum_{\langle i,j \rangle - \text{ближайшие соседи,}} (c_{i,j,\sigma}^\dagger c_{i,j,\sigma} + h.c.) + U \sum_{i,j} n_{i,j,\uparrow} n_{i,j,\downarrow}$$

$c_{i,j,\sigma}^\dagger/c_{i,j,\sigma}$ — операторы рождения и уничтожения электрона на узле с индексом i в продольном направлении, индексом j в поперечном направлении и спином σ , t_{\parallel} — амплитуда туннелирования в продольном направлении, t_{\perp} — амплитуда туннелирования в поперечном направлении, U — потенциал взаимодействия электронов, находящихся на одном узле, а $h.c.$ означает эрмитово сопряжение. В расчетах принято $t_{\parallel,\sigma} = t_{\perp,\sigma}$ и $t_{\uparrow} > t_{\downarrow}$ (фермионы со спином вниз тяжелые), а взаимодействие притягивающее ($U < 0$). Расчет сверхпроводящие корреляции на лестничных решетках рассчитывались методом DMRG (**D**ensity **M**atrix **R**enormalization **G**roup)[11]. Использована реализация метода в пакете iTensor и его интерфейс для языка C++[12].

Расчеты

В предыдущей работе [13] было показано, что с ростом количества направляющих в лестничной решетке, энергия связи тримеров падает. Для строго одномерного случая именно наличие связанных состояний трех и более фермионов подавляет сверх-

проводящие корреляции [14]. В работе производится расчет волновой функции основного состояния при фиксированном количестве фермионов со спинами вверх и вниз, а затем по полученной функции вычисляется сверхпроводящий коррелятор для каждой пары узлов i, j вида:

$$\rho_{i,j} = \langle \psi_{i\uparrow} \psi_{i\downarrow} \psi_{j\uparrow}^\dagger \psi_{j\downarrow}^\dagger \rangle$$

Изменяемыми параметрами для рассматриваемых систем являлись ширина лестницы W , а также плотность носителей заряда с разными направлениями спина. Для одномерного случая, корреляции такого вида будут экспоненциально убывать при соизмеримых в масштабе 2 к 1 количествах фермионов (тяжелых должно быть вдвое больше). При отходе от соизмеримости в любую сторону подавление сверхпроводимости за счет тримеров исчезает [14]. Для расчетов использованы значения $U = -5t_{\uparrow}$, $t_{\downarrow} = 0.3t_{\uparrow}$ или $t_{\downarrow} = t_{\uparrow}$, $L = 80$

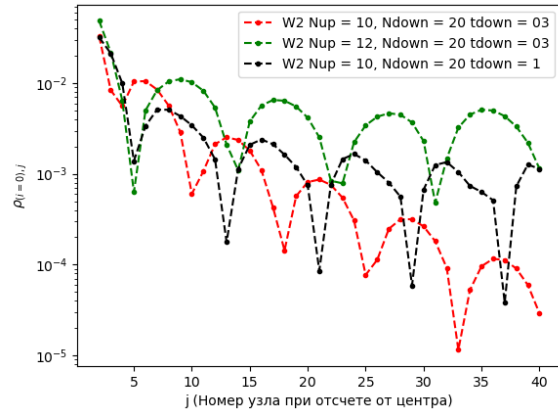


Рис.2. Сверхпроводящие корреляции $\rho_{i=0,j}$ на лестничной решетке при $W=2$. Шаг по оси X отсчитывается от центрального узла вдоль одной из направляющих, по оси Y масштаб логарифмический

На рис.2 приведены зависимости $\rho_{i,j}$ из центра решетки при меняющихся числах заполнения. Невооруженным глазом видно, что значения корреляций от расстояния спадают экспоненциально при соизмеримых плотностях и алгебраически — при несоизмеримых плотностях или $t_{\downarrow} = t_{\uparrow}$. Это означает, что наличие тримеров действительно разрушает дальний порядок при соизмеримых плотностях носителей не только в одномерном, но и в квазиодномерном случае.

Для случая $W=3$ результат уже менее очевиден.

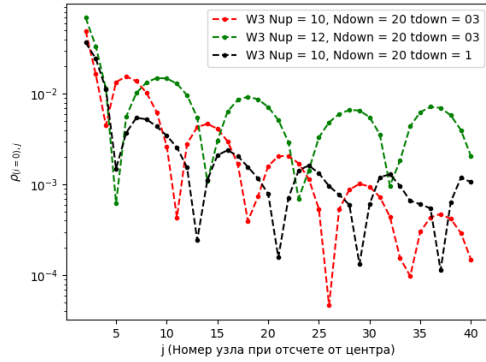


Рис.3. Сверхпроводящие корреляции $\rho_{i=0,j}$ на лестничной решетке при $W=2$, Шаг по оси X отсчитывается от центрального узла вдоль центральной направляющей, по оси Y масштаб логарифмический

На рис.3 видно, что в случае неравных амплитуд туннелирования скорость убывания корреляций при соизмеримых плотностях фермионов выше, чем при несоизмеримых, что должно свидетельствовать в пользу влияния тримеров. Однако если сравнить скорость убывания при одинаковых плотностях и различных соотношениях $t_{\downarrow}/t_{\uparrow}$, точный вывод о том, где скорость убывания больше, сделать нельзя. Это означает, что если явное влияние связанных состояний на сверхпроводимость и есть, то оно менее заметно, чем на меньшем количестве направляющих. Чтобы удостовериться в этом выводе были рассмотрены корреляции на лестничных решетках с 4 и 5 направляющими.

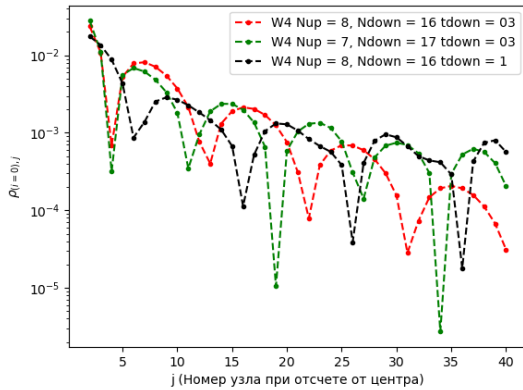


Рис.4 Сверхпроводящие корреляции $\rho_{i=0,j}$ на лестничной решетке при $W=4$. Шаг по оси X отсчитывается от центрального узла вдоль одной из центральных направляющих, по оси Y масштаб логарифмический

На рис.4 видно, что разница в скорости убывания корреляций стала еще меньше, чем для случая $W = 3$, что, в целом согласуется, с выводом [13] о том, что с ростом количества направляющих (переходом от 1D режима к 2D режиму), влияние тримеров спадает.

Кроме вышесказанного стоит отметить, что влияние тримеров уменьшается не только с шириной

лестничной решетки, но и с уровнем ее заполнения. Для этого аналогично работе [5] можно рассмотреть зависимость тримерной щели от заполненности решетки при соизмеримых плотностях. Тримерная щель вычисляется следующим образом:

$$\Delta_{trimer} = -(E(N_{\uparrow} + 1, N_{\downarrow} + 2) + E(N_{\uparrow}, N_{\downarrow}) - E(N_{\uparrow} + 1, N_{\downarrow} + 1) - E(N_{\uparrow}, N_{\downarrow} + 1))$$

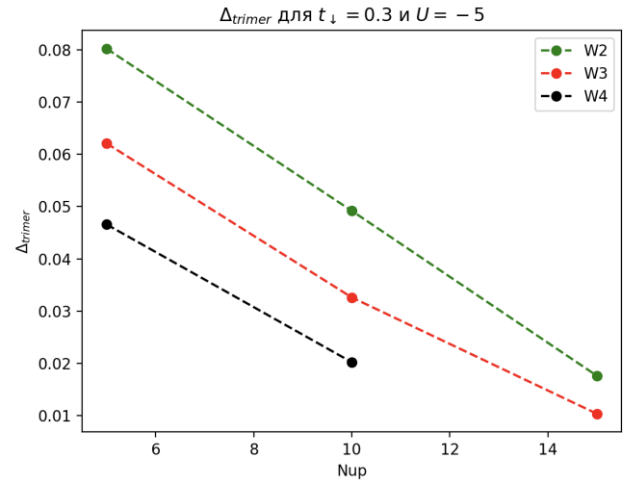


Рис.5. Тримерная щель для решеток с $W=2,3,4$ и $2n_{\uparrow} = n_{\downarrow}$

Заключение

Получены графики сверхпроводящих корреляций при соизмеримых и несоизмеримых плотностях носителей заряда в случае с одинаковыми и различными амплитудами туннелирования для лестниц с 2-4 направляющими в модели Хаббарда. Качественно подтвержден предыдущий вывод о том, что с переходом от строго одномерного режима к квазиодномерному влияние связанных состояний 3 частиц уменьшается. Кроме того, показано, что с ростом чисел заполнения, тримерная щель уменьшается, что говорит о влиянии тримеров только при малых числах заполнения. Исследование выполнено с использованием суперкомпьютерного комплекса НИУ ВШЭ.

Список литературы

1. M. C. Reville, J. A. Fry, B. A. Olsen and R. G. Hulet, 1D to 3D Crossover of a Spin-Imbalanced Fermi Gas, Phys. Rev. Lett. 117 235301 (2016)
2. Fulde, P. & Ferrell, R. A. Superconductivity in a Strong Spin-Exchange Field. Phys. Rev. 135, A550–A563 (1964)
3. Larkin, A. I. and Ovchinnikov, Y. N. Inhomogeneous state of superconductors. Sov. Phys. JETP 20, 762-769 (1965)
4. Partridge et al., Deformation of a trapped Fermi gas with unequal spin populations, Phys. Rev. Lett. 97, 190407 (2006)
5. G. Orso Attractive Fermi Gases with Unequal Spin Populations in Highly Elongated Traps Phys. Rev. Lett. 98, 070402 (2007)

6. A. Feiguin and F. Heidrich-Meisner, Pairing States of a Polarized Fermi gas trapped in a one-dimensional optical lattice, *Phys. Rev. B* 76, 220508(R) (2007)
7. A.Potapova, I.Pilé, Tian-Cheng Yi, R.Mondaini and E.Burovski, Dimensional crossover on multileg attractive-U Hubbard ladders, *Phys. Rev. A* 107, 053301 (2023)
8. G. Orso, L. P. Pitaevskii, and S. Stringari, Equilibrium and dynamics of a trapped superfluid Fermi gas with unequal masses, *Phys. Rev. A* 77, 033611 (2008)
9. Huang B., Fritsche I., Lous R. S., Baroni C., Walraven J. T. M., Kirilov E. and Grimm R., Breathing mode of a Bose-Einstein condensate repulsively interacting with a fermionic reservoir, *Phys. Rev. A*, 99, 041602 (2019)
10. O. Mandel et al., Coherent Transport of Neutral Atoms in Spin-Dependent Optical Lattice Potentials, *Phys. Rev. Lett.* 91, 010407 (2003)
11. U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Annals of Physics*, Volume 326, Issue 1, 96-192 (2011)
12. Fishman, S. R. White, E. M. Stoudenmire, The ITensor Software Library for Tensor Network Calculations <https://arxiv.org/abs/2007.14822> (2020)
13. Пиле, Я. Э. Связанные состояния в модели Хаббарда на лестничных решетках / Я. Э. Пиле // Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов имени Е.В. Арменского, Москва, 27 февраля – 07 марта 2023 года. – Москва: Московский институт электроники и математики НИУ ВШЭ, – С. 56-59 (2023)
14. G.Orso, E.Burovski, and T.Jolicoeur. Luttinger Liquid of Trimers in Fermi Gases with Unequal Masses, *Phys. Rev. Lett.* 104, 065301 (2010)

**АНАЛИЗ И ОЦЕНКА
ПРОИЗВОДИТЕЛЬНОСТИ
СУПЕРКОМПЬЮТЕРА “CHARISMA” И ЕГО
КОМПОНЕНТОВ ДЛЯ ПЕРВОПРИНЦИПНОГО
МОДЕЛИРОВАНИЯ МАТЕРИАЛОВ**

*Недомолкин И.Э., Коников М.П.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ*

Аннотация

Цель данного исследования состоит в проведении анализа и оценки производительности суперкомпьютера “CHARISMa” на популярных научных задачах. В работе исследуется производительность программного пакета первопринципной молекулярной динамики Quantum ESPRESSO и CP2K на различных узлах суперкомпьютера CHARISMa.

Введение

В современной эпохе вычислительной техники суперкомпьютерные системы, также известные как системы высокопроизводительных вычислений (HPC), играют ключевую роль в проведении научных и инженерных исследований. Их значимость постоянно возрастает в связи с непрерывным расширением области их применения. Они используются для решения сложных задач в различных областях науки, включая квантовую химию и физику [1].

Один из наиболее известных рейтингов суперкомпьютеров основан на тестах LINPACK [2]. Данный бенчмарк измеряет как быстро тестируемое устройство находит решение плотной системы линейных уравнений [3]. Это довольно эффективный способ так как данная задача является весьма распространённым типом расчётов. Однако стоит отметить, что большинство суперкомпьютеров специализируются и заточиваются под решения задач подобного типа, что может исказить результаты, делая их менее репрезентативными в вопросах прикладах задач. [4]

По этой причине использование реальных научных пакетов является более эффективным и точным методом чтобы узнать производительность в конкретном классе научных задач. [5]

Основной задачей данной работы является проведение собственных исследований эффективности суперкомпьютеров и их компонентов при работе с задачами квантовой химии, и составление рейтинга суперкомпьютеров и их компонентов на основе полученных результатов.

Методы и подходы

Был выбран перечень самых популярных программных пакетов для проведения расчётов в области квантовой химии, после чего осуществлён поиск, по ключевым словам, чтобы узнать, как часто тот или иной программный пакет использовался в научных статьях и различных публикациях (Таблица 1). Благодарим этому анализу можно будет сделать результаты нашей работы более репрезентативными для большей группы пользователей. В результате был выбран пакет Quantum Espresso [6][7][8] и CP2K [9]. Они являются достаточно распространёнными в научной сфере, распространяются бесплатно, имеют открытый исходный код и реализуют самые популярные методологии расчётов.

Таблица 1. Частота упоминания программных пакетов

| Пакет | Simulations | Benchmark | Quantum chemistry | All |
|----------|-------------|-----------|-------------------|-------|
| VASP | 71 400 | 12 800 | 51 600 | 9 710 |
| CPMD | 8 500 | 1 840 | 6 390 | 1 540 |
| CP2K | 7 720 | 2 080 | 5 560 | 1 890 |
| TeraChem | 918 | 383 | 1 020 | 358 |

| | | | | |
|------------------|--------|--------|--------|-------|
| Abinit | 7 610 | 1 410 | 5 720 | 1 010 |
| Quantum ESPRESSO | 25 500 | 5 230 | 26 300 | 3 730 |
| Wien2K | 10 400 | 1 010 | 9 340 | 651 |
| GAMESS | 28 100 | 23 500 | 49 900 | 3 690 |
| GPAW | 2 450 | 721 | 18 700 | 618 |
| PFARM | 198 | 82 | 34 | 29 |
| Orca | 29 500 | 12 700 | 18 600 | 4 110 |
| BigDFT | 533 | 247 | 402 | 212 |

Quantum ESPRESSO разработан на языках Fortran и C. В нём используются библиотеки: BLAS, LAPACK, FFT, CUDA библиотеки и другие [10].

Один из методов, используемых в пакете Quantum ESPRESSO при моделировании – это метод функционала плотности, такой же метод и его вариации используются и в CP2K. При расчёте задачи происходит решение дифференциального уравнения второго порядка. Это происходит путём решения линейной системы [11], которая в свою очередь не обязательно должна быть плотной.

У QE существует множество бенчмарков, но самым популярным является AUSURF112 [12]. Он используется для расчета электронной структуры поверхности кластера золота. Этот бенчмарк определяет расстояние между двумя слоями атомов металлического кластера, определяет возможные узконаправленные переходы, определяет свойство и структуру поверхности. Он входит в перечень тестовых задач, на которые указывают сами разработчики программного пакета Quantum ESPRESSO.

Для анализа производительности пакетом CP2K был выбран бенчмарк H2O-128[13], как типовая эталонная задача в молекулярной динамике. В первую очередь исследуется молекулярная динамика воды, используя приближение Борна-Оппенгеймера и теории функционала плотности. Бенчмарк H2O-128.inp исследует систему из 128 молекул воды (384 атома, 1'024 электронов).

Работа с Quantum ESPRESSO и CP2K

Тестирование проводилось на суперкомпьютере “сHARISMa” [14]. Ниже приведены характеристики его узлов. Так как запуски проводятся без использования графических процессоров основной вклад в производительность делают процессоры, поэтому сравнительные характеристики считаются для них, а не для всего узла. То есть нормировка на цену производится не с использованием себестоимости конкретного узла, а только на его процессор.

Таблица 2. Характеристики узлов Суперкомпьютера

| | | | | |
|---------------------------------------|----------------|----------------|----------------|----------------|
| Типы узлов сHARISMa | Type a | Type b | Type c | Type d |
| Наименование вычислительно-устройства | 2 x Intel Xeon | 2 x Intel Xeon | 2 x Intel Xeon | 2 x Intel Xeon |

| | | | | |
|---|-------------------------|-------------------------|-------------------------|-----------------------|
| | Gold 6152 2.1 – 3.7 ГГц | Gold 6152 2.1 – 3.7 ГГц | Gold 6240 R 2.4 – 4 ГГц | Gold 6248 R 3 – 4 ГГц |
| Количество ядер | 2*22 ядер | 2*22 ядер | 2*24 ядер | 2*24 ядер |
| Цена CPU | 3655\$ | 3655\$ | 2592\$ | 3178\$ |
| RAM | 768 ГБ | 1536 ГБ | 768 ГБ | 376 ГБ |
| Энергопотребление | 140 Ватт | 140 Ватт | 165 Ватт | 205 Ватт |
| Пиковая производительность | 5.2 TFlops | 5.2 TFlops | 6.1 TFlops | 6.1 TFlops |
| Пропускная способность STREAM на всем CPU | 74206 МБ/с | 68413 МБ/с | 79835 МБ/с | 78481 МБ/с |

На этих узлах была проведена серия запусков бенчмарков. Запуски производились на 1, 2, 4, 8, 16, 22 и 24. Мpi-процессах. В таблице 3 и таблицу 4 приведены времена выполнения бенчмарков на всех ядрах одного процессора. Строка all соответствует полному использованию процессора. Стоит отметить, что часть узлов использует 22 ядерные процессоры (узлы А и В), а другие (узлы С и D) 24.

Таблица 3. Время выполнения бенчмарка QE

| | | | | |
|-----|--------|--------|--------|--------|
| MPI | Узел А | Узел В | Узел С | Узел D |
| All | 180 | 184 | 158 | 160 |

Таблица 4. Время выполнения бенчмарка CP2K

| | | | | |
|-----|--------|--------|--------|--------|
| MPI | Узел А | Узел В | Узел С | Узел D |
| All | 205 | 210 | 148 | 145 |

Данные были получены в результате 15–20 запусков бенчмарка на каждом выбранном значении mpi-процессов и выборе наилучшего из них. В результате тестов узлы С и D показали лучшее время по сравнению с узлами А и В.

На основе полученных данных бенчмарков и профилировщика были получены значения Gflops и различные нормировки для каждого узла (таблица 5, таблица 6). Лучшие показатели, согласно CP2K, у узла D. Пакет QE напротив, показал лучшую производительность на узле С. Оба бенчмарка повреждают, что узлы А и В уступают двум другим по производительности.

Таблица 5. Нормировки QE

| | Узел А | Узел В | Узел С | Узел D |
|-----------------------------------|------------|------------|------------|------------|
| Производительность | 696 GFlops | 681 GFlops | 793 GFlops | 783 GFlops |
| Нормирование на цену | 0.190 | 0.186 | 0.217 | 0.214 |
| Нормирование на энергопотребление | 4.972 | 4.864 | 5.665 | 5.594 |

Таблица 6. Нормировки CP2K

| | Узел А | Узел В | Узел С | Узел D |
|-----------------------------------|--------------|--------------|--------------|--------------|
| Производительность | 628.5 GFlops | 613.5 GFlops | 870.5 GFlops | 888.6 GFlops |
| Нормирование на цену | 0.172 | 0.167 | 0.335 | 0.279 |
| Нормирование на энергопотребление | 4.448 | 4.382 | 5.276 | 4.334 |

Следующим важным показателем является производительность, нормированная на стоимость. С помощью этой нормировки можно подобрать тот процессор, что более выгоден с экономической точки зрения для решения задач квантовой химии. Чем больше значение, тем выгоднее приобретение с точки зрения производительности за цену. Согласно данной нормировке, процессор Узла С наиболее предпочтителен к покупке, в этом сходятся оба бенчмарка

Другим важным показателем является энергопотребление задачи. Энергоэффективность – это сколько единиц вычислительной мощности (GFlops) получает пользователь за ватт (W). Расчёты проводились на основе теоретического энергопотребления процессора. В обоих тестах лучшие результаты показал Узел С. Данная нормировка демонстрирует эксплуатация какого процессора обойдётся дешевле.

Заключение

В ходе исследования был проведён обзор методов вычисления производительности, научных программных пакетов и бенчмарков в рамках квантовой химии. Выбраны два популярных научных пакета для определения производительности вычислительных систем в рамках задач первопринципного моделирования материалов.

В ходе параллельного многочисленного тестирования было найдено время выполнения эталонных задач при разном количестве вычислительных потоков на различных узлах, на основе которых был проведен анализ с нормированием результатов и определена производительность узлов суперкомпьютера «сHARISMa». Также было выявлено, что в результате проведённых тестов наилучшие показатели имеет кластер С, с процессором Intel Xeon Gold 6240R.

Список литературы

1. Guest, Martyn F., et al. "High-performance computing in chemistry: NW Chem." *Future Generation Computer Systems* 12.4 (1996): 273-289.
2. TOP500, <https://www.top500.org/>.
3. Dongarra J. J. The linpack benchmark: An explanation //International Conference on Supercomputing. – Berlin, Heidelberg : Springer Berlin Heidelberg, 1987. – С. 456-474.
4. Sayeed, Mohamed, et al. "Measuring high-performance computing with real applications." *Computing in Science & Engineering* 10.4 (2008): 60-70.
5. Dongarra, J., Heroux, M.A. and Luszczek, P., 2016. A new metric for ranking high-performance computing systems. *National Science Review*, 3(1), pp.30-35.
6. Quantum ESPRESSO, <https://www.quantum-espresso.org/>.
7. Scandolo S. et al. First-principles codes for computational crystallography in the Quantum-ESPRESSO package //Zeitschrift für Kristallographie-Crystalline Materials. – 2005. – Т. 220. – №. 5-6. – С. 574-579.
8. Giannozzi P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials //Journal of physics: Condensed matter. – 2009. – Т. 21. – №. 39. – С. 395502.
9. CP2K, <https://www.cp2k.org/>
10. User's Guide for QUANTUM ESPRESSO (v.7.2), https://www.quantum-espresso.org/Doc/user_guide/.
11. di Cataldo S. A quick introduction to Quantum Espresso. – 2019.
12. Quantum ESPRESSO benchmarks, <https://www.quantum-espresso.org/benchmarks/>.
13. H2O-128, <https://github.com/cp2k/cp2k/tree/master/benchmarks/QS>.
14. HSE University HPC Cluster "сHARISMa", <https://hpc.hse.ru/en/hardware/hpc-cluster/>.

ОПТИМИЗАЦИЯ СЕТКИ КРЕДИТНЫХ
ПРЕДЛОЖЕНИЙ

Рахмаева Е.М., Иванов Г.Д.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ

Аннотация

В работе рассматриваются эвристические методы оптимизации как инструмент поиска наилучшего кредитного предложения (КП). Цель исследования – разработать методологию выделения оптимальной подсетки из крупной сетки КП. Данное исследование проводится в рамках проекта оптимизации процессов кредитования Альфа-Банка.

Введение

Ежедневно в банки поступают тысячи заявок на потребительские кредиты. Перед банком ставится задача определения наиболее выгодного для него предложения. С одной стороны, клиент, подавший заявку, хочет получать максимально выгодные условия по кредиту: большую сумму по низкой ставке на длительный срок, и от предложения банка зависит вероятность взятия кредита. С другой стороны, банк, одобряющий заявку, желает получить максимальную прибыль и удовлетворить свои целевые метрики/функции (TF).

Текущий подход банка заключается в рассмотрении малой сетки - перебора комбинаций небольшого числа параметров: сумма кредита w_1 (руб., до 10 вариантов), ставка w_2 (% , до 5 вариантов), срок кредитования w_3 (мес., 3 варианта), кредитная история w_4 (2 варианта), верификация w_5 - звонок клиенту для подтверждения личности и уточнения цели взятия кредита (2 варианта). Финальное решение (выбор узла сетки – КП) принимается с помощью внутренних моделей банка, которые оценивают доходность сделки (EVA) и вероятность взятия кредита клиентом (TR). Банк понимает, что текущая макросетка КП пропускает много потенциально оптимальных узлов и не учитывает возможные особенности поведения TF для разных клиентов.

Для решения поставленной задачи в данной работе производилось исследование трех основных метаэвристических алгоритмов: генетический алгоритм (GA), алгоритм роевого интеллекта (PS), гравитационный алгоритм (GS).

Задача оптимизации

Для каждого клиента x , подавшего заявку или участвующего в кампании, ставится задача

$$\begin{cases} EVA(x, w) * TR(x, w) \rightarrow max, \\ w_1 \in \{50.000, 55.000, \dots, 7.000.000\}, \\ w_2 \in \{10.00, 10.01, \dots, 30.00\}, \\ w_3 \in \{12, 24, \dots, 60\}, \\ w_4 \in \{0,1\}, w_5 \in \{0,1\} \end{cases} \quad (1)$$

оптимизации (1), где $EVA(x, w)$ - доходность сделки; $TR(x, w)$ - вероятность взятия кредита. Не вдаваясь в подробности построения EVA/TR моделей, скажем лишь, что при исследовании модели выступали как «черные ящики», которые, принимая на вход клиента x и вектор параметров w , выдают оценки.

Подход к решению задачи

Решение задачи (1) по базе X (для каждого клиента в отдельности) крайне затруднительно. К тому же, размер полной сетки – всевозможные комбинации параметров (~55 миллионов) очень велик, и вычислить модельные оценки EVA и TR для каждого узла сетки хотя бы по одному клиенту – уже ресурсоемкая задача. Чтобы обойти эту трудность, банк в настоящее время строит модель кластеризации клиентов, учитывающую их ключевые характеристики и, главным образом, поведение функций EVA и TR . Утверждается, что оптимизация по полной сетке базы X может быть сведена к оптимизации сетки по кластерам $X_j, X = \cup X_j$. А именно, оптимизацию по полной сетке для каждого клиента кластера j можно разбить на два шага:

- 1) оптимизация полной сетки по «центровому» представителю кластера X_j с выделением «суженной подсетки»,
- 2) оптимизация по подсетке для каждого представителя x кластера X_j .

Суженная подсетка включает в себя оптимальные точки центрального представителя с некоторыми окрестностями – на случай, если в кластере находится клиент со слегка измененным поведением TF.

Таким образом, имея для каждого кластера суженные подсетки, банк может выдавать быстрые решения по КП для любого клиента из кластера – для этого достаточно решить задачу оптимизации на малой сетке. Выделению суженной подсетки и будут посвящены следующие разделы.

Метаэвристические методы оптимизации

Многие вопросы, возникающие в прикладных науках, сводятся к задачам условной оптимизации. Частой трудностью здесь становится нелинейность TF, недифференцируемость, многоэкстремальность, отсутствие аналитического выражения и высокая вычислительная сложность оптимизируемых функций, высокая размерность пространства поиска и т.д. Для решения указанных проблем в 80-х годах начали интенсивно разрабатывать стохастические поиско-

вые методы оптимизации. Мы рассматриваем класс *метаэвристических* алгоритмов. В основном, работа любого такого алгоритма начинается со случайной инициализации N точек (популяция первого поколения) в пространстве решений, а затем по каким-то правилам производятся обновления этих точек (2-ое, 3-е, ... поколения). При этом разные методы обладают разными специальными параметрами, которые влияют на качество полученного результата и скорость сходимости алгоритма [1].

Написано немало работ, посвященных модификациям классических метаэвристических алгоритмов [2], созданию новых методов [3], применению существующих алгоритмов в различных прикладных задачах, в том числе в банковском секторе. Среди последних: оптимизация кредитных решений банка с помощью GA [4], решение задачи скоринга клиентов нейронными сетями, оптимизация весов которых реализуется с помощью PS [5] и другие. Однако большинство статей ориентированы либо на оценку кредитоспособности [4] – чтобы определить, имеет ли клиент, обратившийся с заявлением, право на получение требуемого кредита, либо на выбор портфеля, направленный на выбор оптимальных резервов, которые максимизируют прибыль клиентов.

Целью нашей работы является реализация, настройка и выбор лучшего эвристического метода, который обеспечит наиболее точную и быструю (за наименьшее количество поколений и с наименьшим размером популяции) сходимость к реальному максимуму. Этот выбор осуществляется на основании исследования зависимостей между настраиваемыми параметрами алгоритмов и значениями TF. А именно, из-за присутствия случайности в работе методов, для разных значений исследуемого параметра производятся повторные прогоны с фиксированием (на основе best practices статей) остальных параметров и сравнивается усредненное по числу прогонов максимальное (из N) и среднее (по N) значения TF последнего поколения (см. рис.1). Эффективность полученных значений параметров проверяется на нескольких кластерах клиентов.

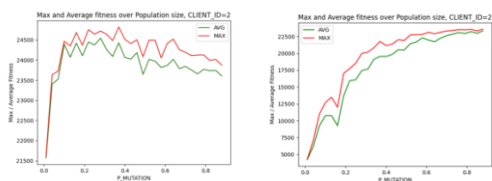


Рис.1. GA: зависимость TF от вероятности мутации при рулетке/панмиксии и смешанном кроссовере

Генетический алгоритм

GA имитирует генетическое наследование и естественный отбор. TF здесь эквивалентна природному понятию приспособленности организмов. Чем «лучше» потомок, тем выше его приспособленность. Вектор параметров при этом называют хромосомой (представитель потомства), а каждую координату вектора геном [6]. Хромосома в нашем

случае – это набор из пяти чисел $w_1- w_5$, каждое из этих чисел – ген.

В данной работе рассматривается GA с вещественным кодированием, включающий в себя шесть основных этапов: формирование начальной популяции, селекцию (отбор представителей популяции для скрещивания), скрещивание (кроссовер), мутацию, формирование новой популяции и проверка условий окончания алгоритма. Мы работаем с селекцией методами панмиксии и рулетки, бинарным и смешанным кроссоверами, нормально распределенной мутацией. Таким образом, нужно выбрать вид селекции и кроссовера, а также определить вероятности кроссовера и мутации, размер популяции и количество поколений.

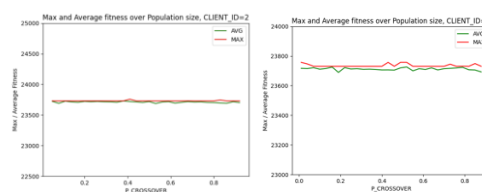


Рис.2. GA: зависимость TF от вероятности кроссовера при панмиксии и смешанном / бинарном кроссовере

Проведя ряд исследований, можно сделать вывод, что для GA не важен метод и вероятность кроссовера (см. рис.2), а также размер популяции (best practices 20 достаточно, увеличение не влияет на TF). Незначительное влияние оказывает метод селекции, существенное влияние оказывается вероятностью мутации (см. рис.1).

В рамках нашего эксперимента лучший результат получен для GA с панмиксией с параметрами: размер популяции 20, количество поколений 30, вероятность мутации 0.8, вероятность кроссовера 0.3.

Рой частиц

Алгоритм имитирует поведение роя. Вектор параметров рассматривается как частица. Каждая частица обладает скоростью, которая со временем изменяется таким образом, чтобы оказаться в наиболее выгодной области [7]. А именно, частица помнит свои местоположение, связанное с наилучшим решением (пригодностью), достигнутом к текущему моменту (*pbest*), и общее наилучшее решение в популяции, полученное к текущему моменту (*gbest*).

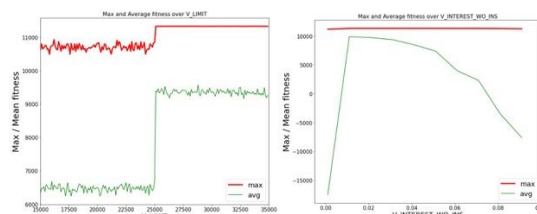


Рис.3. PS: зависимость TF от максимальной скорости по лимиту и ставке

Скорость частицы меняется по направлению к местоположениям ее $pbest$ и $gbest$ за счет случайного взвешивания направлений к местоположениям $pbest$ и $gbest$. Для предотвращения больших скачков скорость в каждом измерении ограничивается максимальной скоростью V_{max} . Таким образом, в качестве параметров в PS выступают: весовые коэффициенты c_j , ограничение для модуля скорости, ускорение ω , а также размер популяции и количество поколений.

В рамках нашего эксперимента наилучшим оказывается PS с параметрами: размер популяции 40, количество поколений 1, весовые коэффициенты, равные 2, и ускорение 1. При этом в качестве V_{max} для лимита выступает 27.000, для срока 7 и для ставки 0.01 (см. рис.3).

Гравитационный поиск

Ньютоновская гравитация — это тенденция масс ускоряться навстречу друг другу. Согласно закону всемирного тяготения Ньютона, каждая частица притягивает любую другую частицу с помощью «гравитационной силы». Гравитационная сила между двумя частицами прямо пропорциональна произведению их масс и обратно пропорциональна квадрату расстояния между ними. В качестве масс в нашей задаче рассматривается значение TF.

Настройка GS происходит за счет изменения размера популяции, числа поколений, а также значения гравитационной постоянной в момент времени t (выбор $G(t_0)$ и константы β):

$$G(t) = G(t_0) \left(\frac{t_0}{t}\right)^\beta. \quad (2)$$

При реализации алгоритма на основе [8] учитывалась возможность возникновения “нулевых” масс. Для этого алгоритм должен учитывать масштаб и область значений вдоль каждой оси.

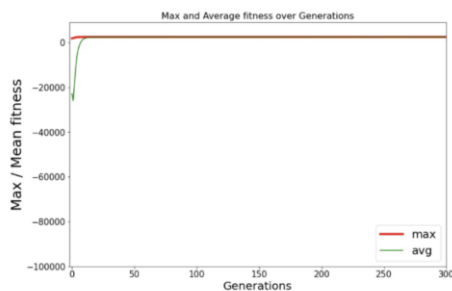


Рис.4. GSA: зависимость TF от количества поколений

В нашей задаче наилучшим оказывается GS с параметрами: $\beta \in [0.01, 0.1]$, G_0 не влияет, количество поколений 20 (см. рис.4), размер популяции 20.

Сравнительный анализ алгоритмов

Для сравнения алгоритмов ставился следующий эксперимент. Фиксируя начальную популяцию, все три алгоритма запускались 100 раз для различ-

ных кластеров клиентов. Далее, сравнивалась разница между средним значением TF текущего поколения (по 100 прогонам) и фактическим максимумом. Результаты сравнения представлены на рис. 5.

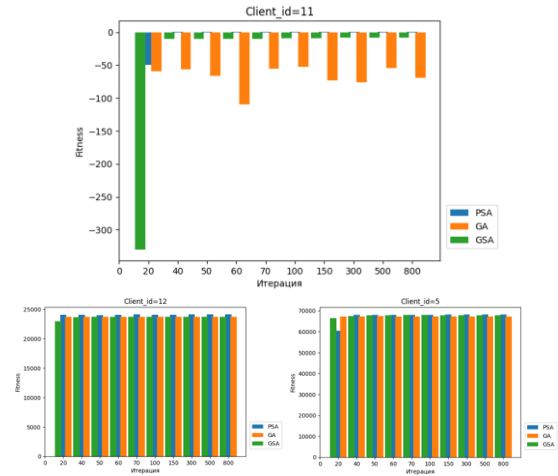


Рис.5. Сравнение эффективности алгоритмов: истинные $\max TF - 0.8, 25794.46, 70395.57$

Для некоторых кластеров клиентов GA испытывает сложности со сходимостью (см. client_id=11; предположительно, это связано с выбранным типом мутации). GS сходится во всех случаях примерно к 40-ому поколению, но TF популяции находится ниже максимума (например, client_id=5). PS также сходится к 40-й итерации и TF популяции оказывается максимальным среди всех трех алгоритмов (в большинстве случаев, полученное значение TF совпадает с реальным максимумом).

Таким образом, для оптимизации сетки кредитных предложений произвольного клиента лучше всего подходит алгоритм роя частиц. Оптимальными параметрами для его работы применительно к поставленной задаче и предоставленным данным являются: размер популяции 40, количество поколений 40, весовыми коэффициентами, равными 2, ускорением 1. При этом в качестве V_{max} для лимита выступает 27.000, для срока 7 и для ставки 0.01.

Список литературы

1. А.В.Пантелеев, Д.В.Мельницкий, Е.А.Алешина. Методы глобальной оптимизации. Метаэвристические стратегии и алгоритмы. Вузовская книга, 2013.
2. Srinivas M, Patnaik L.M. Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms. IEEE Transactions on System, Man, and Cybernetics, 1994.
3. А.П.Карпенко. Популяционные алгоритмы глобальной поисковой оптимизации. Обзор новых малоизвестных алгоритмов. Информационные технологии № 7, 2012.
4. Noura Metawa, Kabir Hassan, and Mohamed Elhoseny. Genetic algorithm based model for optimizing bank lending decisions. Expert Systems with Applications, 2017.
5. A.Ashgari, H.Azgomi, M. Negahi. A new model to improve decision-making indicators of bank customers'

based on neural networks and particle swarm optimization algorithm, 2022.

6. John H. Holland. Genetic algorithms. Scientific America, 1992.

7. C.Russell, Eberhart, and Yuhui Shi. Particle swarm optimization: Developments, applications and resources. In Congress on Evolutionary Computation, 2001.

8. E.Rashedi, H.Nezamabadi-pour, S.Saryazdi. GSA: A Gravitational Search Algorithm. Information Sciences, 2009.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПОЛЕВОЙ ЭМИССИИ ИЗ КАТОДА МАЛЫХ РАЗМЕРОВ

Борисов В.Д.

Национальный исследовательский университет «Высшая школа экономики», аспирантская школа по техническим наукам МИЭМ НИУ ВШЭ

Аннотация

В этой работе мы предлагаем численный метод для решения системы дифференциальных уравнений в частных производных с нарушением непрерывности потока на внутренних поверхностях области определения решения – так называемой задачи Соболева. Этот подход применяется для математического моделирования эмиссии поля из катода малого размера в вакууме.

Введение

На рис.1 представлена область моделирования:

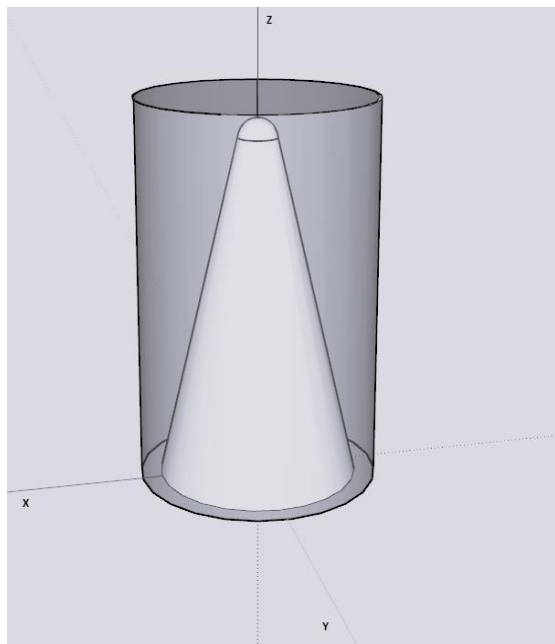


Рис. 1.

Система уравнений, моделирующая полевою эмиссию, имеет следующий вид (1):

$$\begin{cases} \frac{\partial \theta}{\partial t} - \nabla(k_{Si} \nabla \theta) - f = 0 \\ \nabla(\sigma \nabla \psi) = 0 \end{cases} \quad (1)$$

где θ – температура, ψ – потенциал электрического поля, f – источники Джоулева тепла, k_{Si} – коэффициент теплопроводности кремния и σ – коэффициент электропроводности. $\sigma = \sigma_{Si}$ – коэффициент электропроводности кремния (катода), $\sigma = \sigma_V$ – электропроводность вакуума.

Уравнение теплопроводности – первое уравнение в (1), решается внутри катода. Краевые условия для уравнения теплопроводности (2) имеют следующий вид:

$$\begin{aligned} \frac{\partial \theta}{\partial n} \Big|_L = 0, \quad \frac{\partial \theta}{\partial n} \Big|_{z=0, x^2+y^2 \leq R_1^2} = 0, \\ k \frac{\partial \theta}{\partial n} \Big|_\Gamma = \frac{j_{em}}{e} \varepsilon_N \end{aligned} \quad (2)$$

где j_{em} – плотность эмиссионного тока [2], ε_N – средняя энергия эффекта Ноттингама [3], L – боковая поверхность катода, Γ – поверхность эмиссии. Здесь мы пренебрегаем излучением тепла в вакуум, аналогично [1].

Граничные условия для второго уравнения в (1) – уравнения на потенциал, имеют вид (3):

$$\begin{aligned} \psi \Big|_{z=0, x^2+y^2 \leq R^2} = \psi_0, \\ \psi \Big|_{z=h, x^2+y^2 \leq R^2} = 0, \\ \frac{\partial \psi}{\partial n} \Big|_{x^2+y^2=R^2, 0 < z < h} = 0, \\ \left[\sigma \frac{\partial \psi}{\partial n} \right] \Big|_\Gamma = (j_{Si} - j_{em}) \Big|_\Gamma, \end{aligned} \quad (3)$$

где ψ_0 – заданное значение, $[*] \Big|_\Gamma$ – скачок соответствующего значения через поверхность Γ в направлении внешней нормали, j_{Si} – плотность тока внутри катода на поверхности эмиссии.

Решение задачи Соболевского типа

Условие, заданное для поверхности излучения, преобразует обычную краевую задачу в задачу типа Соболева, когда внутри области на некоторой заданной (открытой) поверхности ставится дополнительное условие, аналогичное условию на границе области. Здесь это последнее условие в (3).

Подчеркнем, что задача типа Соболева нарушает условие непрерывности значения $\sigma \frac{\partial \psi}{\partial n}$ на поверхности разрыва коэффициента Γ во втором уравнении в (1), то есть на границе кремний-вакуум.

Численные методы решения задач с разрывными коэффициентами хорошо известны [4], но общепринятых алгоритмов решения задач Соболевского типа не существует. В этой работе мы предлагаем метод сведения рассматриваемой задачи Соболевского типа к обычной задаче с разрывным коэффициентом и условием непрерывности потока на поверхности разрыва коэффициента.

Формально тот же метод может быть применен для решения первого уравнения в (1), то есть уравнения теплопроводности. Предполагая, что коэффи-

циент k_V мал в области вакуума, $k_V \ll k_{Si}$ и задача (1-3) может быть решена во всей цилиндрической области V . В этом случае мы снова получаем задачу Соболевского типа для уравнения теплопроводности в силу условия, моделирующего эффект Ноттингама на поверхности излучения.

Для решения задачи (1-3) используем метод, применяемый для задачи Стефана. Напомним, что на границе раздела в задаче Стефана тепловой поток имеет разрыв, и условие непрерывности потока заменяется условием Стефана $\left[k \frac{\partial \theta}{\partial n} \right]_{\Gamma_t} = \lambda v_n$, где k – это разрывный коэффициент теплопроводности θ – температура, Γ_t – поверхность раздела фаз, n – нормаль к Γ , $\lambda = \text{const}$, v_n – скорость по внешней нормали от Γ .

Идея регуляризации состоит в том, чтобы ввести следующее выражение $-\lambda \frac{\partial H(\Gamma_t)}{\partial t}$, где $H(z)$ – функция Хевисайда для скачка на Γ_t . Это выражение в правой части компенсирует скачок теплового потока на Γ_t . Сглаживая функцию Хевисайда, мы получаем приближенную задачу во всей области, которая может быть решена классическими методами. В этой статье мы сделаем то же самое.

Используем определение обобщенного решения для уравнения на потенциал. Полученный интеграл преобразуем по формуле Остроградского-Гаусса. Все слагаемые обратятся в ноль, кроме слагаемого со скачком по внешней нормали через поверхность катода. При чем, как было сказано ранее, поверхность катода состоит из двух частей – боковой поверхности конуса и поверхности полусферы, на которой происходит эмиссия электронов. Следовательно

$$\int_V (\nabla \phi, \sigma \nabla \psi) dV = \int_\Gamma \phi \left[\sigma \frac{\partial \psi}{\partial n} \right] dS = \left(\delta_\Gamma, \phi \left[\sigma \frac{\partial \psi}{\partial n} \right] \right),$$

где δ_Γ – дельта-функция Дирака на поверхности Γ . Подставим полученное выражение в правую часть уравнения для потенциала. Скачок производной описан в последнем условии в (3). Используя закон Ома, получим следующее выражение (4):

$$\nabla(\sigma \nabla \psi) = \delta_\Gamma \left(-\sigma_{Si} \frac{\partial \psi}{\partial n} - j_{em} \right), \quad (4)$$

Псевдостационарное решение

Уравнение (4) – уравнение на потенциал в цилиндрической системе координат. Мы ищем стационарное решение для плоскости, проходящей через ось симметрии. Область разбивается прямоугольной сеткой с равномерным шагом, уравнение решается с помощью явной схемы, определенной на шеститочечном шаблоне [4].

Для наглядности приведем решение упрощенной линейной задачи. Уравнение имеет следующий вид (5):

$$ky'' = 0. \quad (5)$$

Краевые условия для уравнения (5) аналогичны условиям в задаче (1-3):

$$\begin{aligned} y|_{z=0} &= U, \\ y|_{z=2} &= 0, \\ y|_{z=1-} &= y|_{z=1+}, \\ \left[\frac{\partial y}{\partial z} \right]_{z=1} &= \alpha, \end{aligned} \quad (6)$$

где U и α известные значения. Очевидно, что решение будет состоять из двух частей вида $y = az + b$, а именно, решение будет следующим:

$$\begin{aligned} y_1 &= \frac{\alpha - U}{2} z + U, \\ y_2 &= -\frac{\alpha + U}{2} z + \alpha + U. \end{aligned}$$

В зависимости от знака коэффициента при переменной в y_1 возможны два принципиально различных вида решения: в случае, когда коэффициент отрицательный (рис.2) и когда коэффициент положительный (рис.3).

На (рис.2,3) сплошной линией показано решение задачи, вертикальной пунктирной линией показана граница, на которой происходит скачок, горизонтальной штрихпунктирной линией показано U в начале координат.

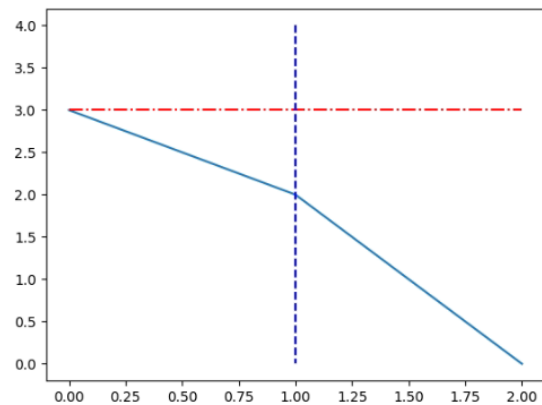


Рис.2. Коэффициент при переменной в y_1 меньше 0

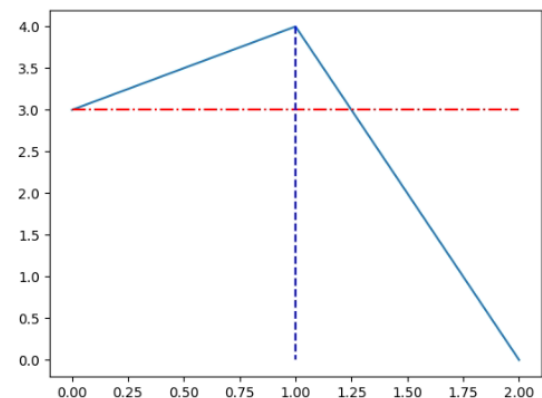


Рис.3. Коэффициент при переменной в y_1 больше 0

В случае (рис.3) на первом отрезке графика потенциал возрастает, а значит движение электронов происходит справа налево, что не соответствует тре-

бованиям исходной задачи. Нас интересует решение, аналогичное (рис.2).

Как показали результаты счета, при поиске стационарного решения уравнения на потенциал (4), сначала мы получаем псевдостационарное решение (рис.4) схожее с (рис.3) после чего оно медленно меняется, стремясь к решению, аналогичному (рис.2).

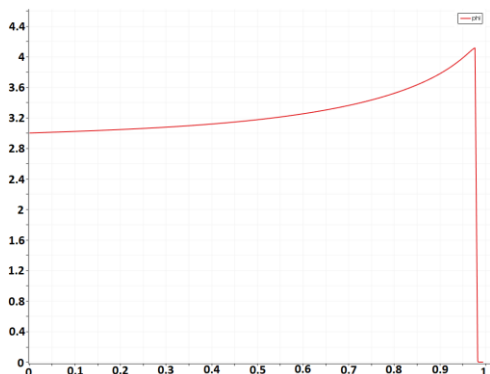


Рис.4. Псевдостационарное решение

Чтобы избежать данной ситуации и значительно сократить поиск стационарного решения, мы правую часть уравнения (4) помещаем под модуль, чтобы в процессе численного счета сохранить у нее знак. Таким образом, мы минуем псевдостационарное решение и задача быстрее сходится к стационару (рис.5).

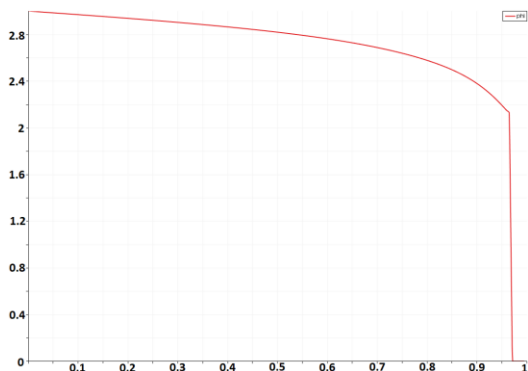


Рис.5. Стационарное решение

Заключение

Таким образом, в данной работе кратко описаны методы, применяемые для поиска стационарного решения уравнения на потенциал численным способом. Описаны способ сведения задачи Соболевского типа к классической задаче дифференциальных уравнений в частных производных с разрывным коэффициентом, а также, способ исключить псевдостационарное решение и сократить вычисления во время моделирования.

Список литературы

1. Danilov V., Gaydukov R., Vadim Kretov, Rudnev V. Modelling of Liquid Nuclei Generation for Field-Emission Silicon Nanocathode // IEEE Transactions on Electron Devices. 2014. Vol. 61. No. 12. P. 4232-4239.

2. Fowler R. H., Nordheim L. Electron emission in intense electric fields //Proceedings of the Royal Society of London. Series~A. – 1928. – Vol.~119, no.~781. – P. 173–181

3. Nottingham W. B. Remarks on Energy losses attending thermionic emission of electrons from metals // Physical Review. – 1941.

4. A.A. Samarsky. Introduction to the theory of difference schemes. // SCIENCE-1971

АНАЛИТИЧЕСКИЕ СВОЙСТВА ОБОБЩЕННОГО УРАВНЕНИЯ ЛЬЕНАРА

Нечитайло В.Г.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ*

Аннотация

Данная работа посвящена исследованию интегрируемости по Лиувиллю обобщенного уравнения Лье́нара, параметризованного многочленами второй и четвертой степеней. Проведен поиск алгебраических и экспоненциальных инвариантов, первых интегралов и интегрирующих множителей уравнения в классе функций Дарбу.

Введение

Обобщенные дифференциальные уравнения Лье́нара имеют вид

$$x_{tt} + h(x_t, x)x_t + g(x) = 0.$$

Они описывают нелинейные осцилляторы с трением, зависящим от скорости, поэтому имеют широкое применение в различных областях науки: от квантовой физики до биологического моделирования. Например, в статье [1] подобные уравнения используются для моделирования нелинейных осцилляторов с переменной плотностью.

Если в уравнении $h(x, x_t) = h(x)$, то оно является классическим уравнением Лье́нара. Таким образом, обобщение заключается в зависимости трения h от скорости x_t .

В данной работе исследуется подкласс уравнений, имеющий вид

$$x_{tt} + (x_t)^2 + f(x)x_t + g(x) = 0,$$

где $f(x)$ и $g(x)$ – многочлены степени m и n соответственно. Легко показать, что уравнение, для которого выполняется соотношение $n \neq 2m$, не интегрируемо по Лиувиллю. В данной работе рассматривается более сложный случай: уравнения с $n = 2m$. Предполагается, что $m = 2$, тогда уравнение может быть параметризовано следующим образом:

$$x_{tt} + (x_t)^2 + (x^2 + \alpha x + \beta)x_t + \frac{1 - \gamma^2}{4}x^4 + \delta x^2 + \sigma x + \varepsilon = 0, \quad (1)$$

где $\alpha, \beta, \gamma, \delta, \sigma, \varepsilon \in \mathbb{C}$.

Уравнению (1) соответствует уравнение первого порядка

$$yy_x + y^2 + (x^2 + \alpha x + \beta)y + \frac{1 - \gamma^2}{4}x^4 + \delta x^2 + \sigma x + \varepsilon = 0. \quad (2)$$

При исследовании уравнения (1) выполняется поиск первых интегралов и интегрирующих множителей в классе функций Дарбу:

$$\prod_{j=1}^N F_j^{d_j}(x, y) \exp(R(x, y)), \quad (3)$$

где $R(x, y) \in \mathbb{C}(x, y), F_j(x, y) \in \mathbb{C}[x, y], d_j \in \mathbb{C}, N \in \mathbb{N}$. Используются следующие обозначения: $\mathbb{C}(x, y)$ – множество рациональных функций относительно переменных x, y с коэффициентами из поля комплексных чисел, $\mathbb{C}[x, y]$ – множество многочленов относительно переменных x, y с коэффициентами из поля комплексных чисел.

Известно, что если первый интеграл или интегрирующий множитель уравнения (1) имеет вид (3), то функции $F_j(x, y)$ являются алгебраическими инвариантами уравнения (2). Для их построения необходимо найти все асимптотические ряды Пуанкаре на бесконечности, удовлетворяющие уравнению (2). Это производится с помощью методов степенной геометрии, описанных в [2]. Алгоритм построения алгебраических инвариантов с помощью рядов Пуанкаре был предложен в работе [3].

Для доказательства существования первых интегралов или интегрирующих множителей искомого вида используются необходимые и достаточные условия, предложенные Зингером [4] и Кристофером [5].

Невырожденный случай

На рис.1 изображен многоугольник Ньютона уравнения (2). $\Gamma_2\Gamma_3$ – единственное ребро, порождающее асимптотики решений на бесконечности. Укороченное уравнение, соответствующее ему, имеет вид

$$y^2 + x^2 y + \frac{1 - \gamma^2}{4} x^4 = 0. \quad (4)$$

Из уравнения можно найти асимптотики:

$$y = \frac{1 \pm \gamma}{2} x^2, \quad x \rightarrow \infty.$$

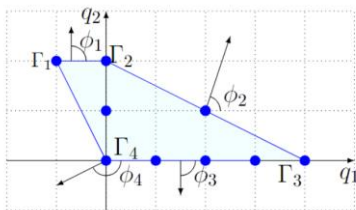


Рис.1. Многоугольник Ньютона уравнения (2)

Случай, когда эти асимптотики не совпадают, называется невырожденным. Для таких уравнений сформулированы необходимые и достаточные условия интегрируемости:

Теорема 1. Уравнение (1) при $\gamma \neq 0$ интегрируемо по Лиувиллю тогда и только тогда, когда оно имеет вид

$$x_{tt} + (x_t)^2 + (x^2 + x + \beta)x_t + \frac{1 - \gamma^2}{4}(x^4 + 2\beta x^2 + x + \beta(\beta + 1)) = 0.$$

Его первый интеграл выражается следующим образом:

$$I(x, x_t) = (x_t - P_1(x))^{d_1} (x_t - P_2(x))^{d_2} \frac{1}{\gamma-1} e^{\frac{2d_1\gamma}{\gamma-1}x},$$

где $P_1(x), P_2(x)$ – полиномиальные решения уравнения (2), имеющие вид

$$P_1(x) = \frac{-1 + \gamma}{2}x^2 + \frac{1 - \gamma}{2}x - \frac{(\beta + 1)(1 - \gamma)}{2},$$

$$P_2(x) = \frac{-1 - \gamma}{2}x^2 + \frac{1 + \gamma}{2}x - \frac{(\beta + 1)(1 + \gamma)}{2}.$$

Пример фазового портрета интегрируемого уравнения представлен на рис.2.

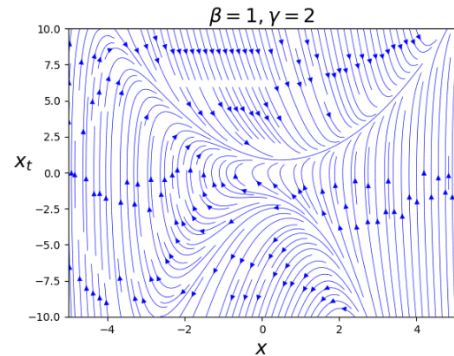


Рис.2. Фазовый портрет интегрируемого случая уравнения (1)

Алгебраическое вырождение

Если $\gamma = 0$, то уравнение (2) можно параметризовать следующим образом:

$$yy_x + y^2 + \left(x^2 + \left(\frac{v^2}{2} + 1\right)x + \beta\right)y + \frac{1}{4}x^4 + \delta x^2 + \sigma x + \varepsilon = 0. \quad (5)$$

В таком случае уравнение на асимптотики (4) имеет кратные корни. Чтобы полностью исследовать асимптотическое поведение решений в окрестности бесконечности, необходимо сделать замену $y(x) = w(x) - \frac{1}{2}x^2$. В результате получим уравнение (6).

$$w_x \left(w - \frac{x^2}{2}\right) + w^2 + \left(\frac{v^2}{2}x + \beta\right)w -$$

$$-\frac{\nu^2}{4}x^3 + \frac{2\delta - \beta}{2}x^2 + \sigma x + \varepsilon = 0. \quad (6)$$

Многоугольник Ньютона, соответствующий уравнению (6) представлен на рис.3.

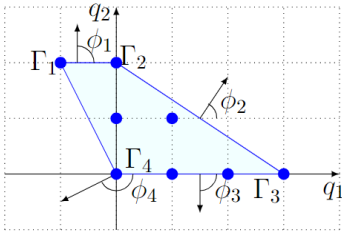


Рис.3. Многоугольник Ньютона уравнения (6)

Все асимптотики решений этого уравнения на бесконечности даны ниже.

$$w(x) = \pm \frac{\nu}{2}x^{\frac{3}{2}}, \quad x \rightarrow \infty. \quad (7)$$

Показатель степени асимптотики нецелый, поэтому уравнение (6), а значит и уравнение (5) не имеет полиномиальных решений. Также, если асимптотики (7) не совпадают, то не выполняются условия существования первого интеграла или интегрирующего множителя вида (3). Из этого следует следующая теорема.

Теорема 2. Уравнение (1) при $\gamma = 0$ и $\nu = \sqrt{2\alpha - 2} \neq 0$ не интегрируемо по Лиувиллю.

Трансцендентное вырождение

Рассматривается частный случай уравнения (5), когда $\nu = 0$. Его можно записать в виде

$$y y_x + y^2 + (x^2 + x + \beta)y + \frac{1}{4}x^4 + \frac{1 + 8\beta - \theta^2}{16}x^2 + \sigma x + \varepsilon = 0. \quad (8)$$

Для того чтобы избежать кратных корней в укороченном уравнении (4), при поиске асимптотик решений уравнений (8) снова необходимо сделать замену $y(x) = w(x) - \frac{1}{2}x^2$. Для функции $w(x)$ можно записать уравнение (9).

$$w_x \left(w - \frac{x^2}{2} \right) + w^2 + \beta w + \frac{2\delta - \beta}{2}x^2 + \sigma x + \varepsilon = 0. \quad (9)$$

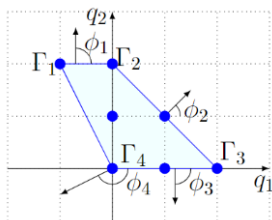


Рис.4. Многоугольник Ньютона уравнения (9)

С помощью многоугольника Ньютона уравнения (рис.4) можно найти все асимптотики решений на бесконечности:

$$w(x) = \frac{1 \pm \theta}{4}x, \quad x \rightarrow \infty. \quad (10)$$

Асимптотики (10) в многоугольнике Ньютона порождает ребро $\Gamma_2\Gamma_3$. Укороченное уравнение этого ребра не является алгебраическим, поэтому легко показать, что при $\theta \in \mathbb{Q}$, асимптотические ряды на бесконечности определены неоднозначно. В противном случае можно доказать следующую теорему.

Теорема 3. Уравнение (1) при $\gamma = 0, \alpha = 1$,

$\theta = \frac{\sqrt{1+8\beta-16\delta}}{4} \notin \mathbb{Q}$ не интегрируемо по Лиувиллю.

Особые случаи

Если в случае трансцендентного вырождения $\theta \in \mathbb{Q}$, то уравнение (6) может иметь более двух рядов с асимптотиками (10), а значит, более сложные первые интегралы и интегрирующие множители. Приведем примеры таких случаев.

Теорема 4. Если уравнение (8) имеет три полиномиальных решения, то соответствующее ему уравнение второго порядка (1) интегрируемо по Лиувиллю и имеет вид

$$x_{tt} + x_t^2 + \left(x^2 + x - \rho^2 - \frac{1}{4} \right) x_t + \frac{1}{4}x^4 - \frac{4\rho^2 + 1}{8}x^2 + \frac{16\rho^4 - 8\rho^2 + 1}{64} = 0, \quad (11)$$

где $\rho \in \mathbb{C}/\{0\}$. Интегрирующий множитель уравнения и соответствующий первый интеграл выражаются следующим образом:

$$M(x, y) = \frac{(y - P_2(x))^{\frac{1}{2\rho}}}{(y - P_1(x))^{\frac{1}{2\rho}}(y - P_3(x))^2},$$

$$I(x, y) = \frac{2(y - P_1(x))^{1-\frac{1}{2\rho}}(y - P_2(x))^{1+\frac{1}{2\rho}}}{y - P_3(x)} - \frac{4\rho^{1-\frac{1}{2\rho}}}{2\rho + 1}(y - P_2(x))^{1+\frac{1}{2\rho}} * F_{21} \left(1 + \frac{1}{2\rho}, \frac{1}{2\rho}, 2 + \frac{1}{2\rho}, \frac{P_2(x) - y}{\rho} \right).$$

Здесь F_{21} – гипергеометрическая функция, а $P_1(x), P_2(x), P_3(x)$ – полиномиальные решения уравнения (8), которые имеют вид

$$P_{1,2}(x) = 1/8(-4x^2 + 4\rho^2 \mp 4\rho + 1), \quad (12)$$

$$P_3(x) = 1/8(-4x^2 + 4x + 4\rho^2 + 1). \quad (13)$$

Теорема 5. Если уравнение класса (1) записывается в виде

$$x_{tt} + (x_t)^2 + \left(x^2 + x - \frac{N^2 + 2N + 2}{4(N+1)^2}\right)y + \\ x_{tt} + (x_t)^2 + \left(x^2 + x - \frac{N^2 + 2N + 2}{4(N+1)^2}\right)y + \\ + \frac{1}{4}x^4 - \frac{N^2 + 2N + 2}{8(N+1)^2}x^2 + \left(\frac{N(N+2)}{8(N+1)^2}\right)^2 = 0,$$

то оно имеет следующий первый интеграл:

$$I(x, y) = (y - P_1(x))^d * \\ * \exp \left[\frac{\sum_{k=0}^N u_k(x)(y - P_1(x))^k}{(y - P_1(x))^N (y - P_3(x))} + h_0 x \right].$$

Здесь полиномы $P_1(x), P_3(x)$ задаются формулами (12), (13) с учетом $\rho = \frac{1}{N+2}$; функция $u_N(x)$ – многочлен степени 2, функции $u_k(x), k = 1, \dots, N-1$ – многочлены степени 1. Явные выражения опущены.

Заключение

В работе исследовалась интегрируемость по Ливиллю обобщенного уравнения Льенара, параметризованного многочленами второй и четвертой степеней. Были сформулированы и доказаны необходимые и достаточные условия интегрируемости в невырожденном случае и в случае алгебраического вырождения. Также доказано, что в случае трансцендентного вырождения уравнение не интегрируемо, если определенный параметр является иррациональным. Были найдены новые примеры интегрируемых уравнений с первыми интегралами в виде функций Дарбу, имеющих экспоненциальные множители с рациональным аргументом.

Данные результаты могут быть применимы в физике при изучении различных нелинейных осцилляторов.

Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ в 2024 году.

Список литературы

1. Ruiz Serván A. Exact solutions to a family of position-dependent mass damped oscillators from variational λ -symmetries / Ruiz Serván A., Muriel Patino M. C. // *Mathematical Methods in the Applied Sciences*. – 2024. – V. 47. – № 2. – P. 891-906.
2. Брюно А. Д. Асимптотики и разложения решений обыкновенного дифференциального уравнения / Брюно А. Д. // *Успехи математических наук*. – 2004. – Т. 59. – №. 3. – С. 31-80.
3. Demina M. V. The method of Puiseux series and invariant algebraic curves / Demina M. V. // *Communications in Contemporary Mathematics*. – 2022. – V. 24. – № 03. – P. 2150007.
4. Singer M. F. Liouvillian first integrals of differential equations / Singer M. F. // *Transactions of the American Mathematical Society*. – 1992. – V. 333. – № 2. – P. 673-688.
5. Christopher C. Liouvillian first integrals of second order polynomial differential equations / Christopher C. // *Electronic Journal of Differential Equations*. – 1999. – V. 1999. – № 49, P. 1-7.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДИНАМИКИ МНЕНИЙ С ОГРАНИЧЕННЫМ ДОВЕРИЕМ С ВНЕШНИМ ВЛИЯНИЕМ

Иваненко О.В., Шаяхметова Д.Т.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ

Аннотация

В данной работе изучаются математические модели динамики мнений с ограниченным доверием с внешним влиянием, основанные на базовой модели Хегсельмана-Краузе (Х-К модель). Предполагается, что на динамику мнений оказывает влияние не только предписанный алгоритм взаимодействия (алгоритм Х-К), но и внешние факторы. Рассматриваются несколько подходов к моделированию влияния внешних факторов. В первой модели предполагается, что с некоторой положительной вероятностью μ мнение агента меняется спонтанно (по некоторому заданному закону распределения), а с вероятностью $1-\mu$ агент действует по классическому алгоритму Х-К. Во второй модели изменение мнения задается линейной выпуклой комбинацией алгоритма Х-К и случайностями двух типов. Для обеих моделей исследуется поведение профиля мнений для разных параметров модели, выявляется влияние параметров случайности на структуру профиля на больших временах.

Введение

Модели динамики мнений представляют собой математический аппарат для изучения процесса формирования мнений в группе взаимодействующих агентов. Модели динамики мнений в настоящее время находят широкое применение как при изучении социальных взаимодействий, в частности эволюции мнений в онлайн-сообществах, так и в технических областях, например, при моделировании движения автономных устройств [1], [2], [3], [4], [5], [6]. Процесс формирования мнений в группе определяется многими факторами, например, графом взаимодействия, алгоритмом взаимодействия, внешними факторами. В частности, структура взаимодействия может быть локальной или глобальной. Под локальным взаимодействием понимается правило, при котором агенты принимают во внимание только своих «ближайших» соседей. При глобальной структуре агенты знают и учитывают информацию всех участников группы.

В данной работе за основу была взята одна из моделей с локальной структурой взаимодействия, которая получила большое распространение, а именно, модель с ограниченным доверием Хегсельмана-Краузе. В моделях с ограниченным доверием предполагается, что на формирование мнения оказывают влияние только те агенты, мнения которых не сильно отличаются от мнения данного агента. Для модели Х-К возможны две ситуации: мнения всех агентов одинаковы (говорим, что система пришла к консен-

сусу), либо все агенты разбились на несколько групп (кластеров), в каждом кластере мнения агентов совпадают. Наличие консенсуса или кластеров зависит от многих факторов: типа распределения начальных мнений агентов, степени их открытости (характеризуется радиусом доверия), числа агентов в группе [7]. Классическая модель Х-К задается детерминированным алгоритмом обновления мнений, случайными в данной модели рассматривались лишь начальные данные. В дальнейшем во многих работах стали появляться стохастические обобщения модели, так как при передаче информации могут быть ошибки, вызванные неточностью измерений или искажениями при передаче информации, а также агенты могут быть подвержены каким-либо внешним воздействиям. Основные способы учета случайных воздействий, следующие: ввод в алгоритм обновления мнений аддитивного случайного шума [8]-[9] и рассмотрение спонтанных изменений мнений [10].

В данной работе мы строим и исследуем стохастические модели на базе классической Х-К модели, которые отличаются от существующих моделей структурой внешнего случайного воздействия, а именно, мы предполагаем, что имеется два типа воздействия: условно положительное и условно отрицательное, которые можно интерпретировать как наличие двух противоположных источников информации. Мы исследуем поведение профиля мнений для разных параметров модели, анализируем влияние параметров случайности на структуру профиля на больших временах.

Модель Хегсельмана-Краузе

Имеется группа из n экспертов (или агентов), через $x_i(t)$ обозначим мнение агента i в момент времени t . Мнения агентов формируют вектор (или профиль) мнений $\vec{x}(t)$. Предполагается, что у каждого эксперта (агента) имеется уровень толерантности (доверия) ε . Тогда на агента i оказывают влияние мнения только тех агентов, которые попали в область его доверия $\mathcal{N}_i(t) = \{j: |x_j(t) - x_i(t)| \leq \varepsilon\}$. Алгоритм обновления мнений имеет следующий вид: мнение агента усредняется по мнениям всех его соседей (включая его самого)

$$x_i(t+1) = \frac{1}{|\mathcal{N}_i(t)|} \sum_{j \in \mathcal{N}_i(t)} x_j(t)$$

Для модели Х-К, в отличие от многих других моделей динамики мнений, справедливо утверждение: для любого начального профиля $x(0)$ система попадает в неподвижную конфигурацию системы за конечное число шагов. Неподвижными конфигурациями являются консенсус или кластеры.

Модель 1 с ограниченным доверием со случайными скачками мнений

Добавим в модель Н-К случайное воздействие: разрешим агенту менять свое мнение в соответствии с некоторым законом распределения. Алгоритм изменения мнений примет вид:

$$x_i(t+1) = \begin{cases} \xi_i(t+1), & \text{если } \eta_i(t+1) = 1 \\ \frac{1}{|\mathcal{N}_i(t)|} \sum_{j \in \mathcal{N}_i(t)} x_j(t), & \text{если } \eta_i(t+1) = 0 \end{cases}$$

$\eta_i(t)$ – независимые случайные величины с распределением Бернулли:

$$\eta_i(t) = \begin{cases} 1, & \text{с вероятностью } \mu \\ 0, & \text{с вероятностью } 1 - \mu \end{cases}$$

$\xi_i(t)$ – независимые случайные величины (н.с.в.), распределение которых с вероятностью β_1 равномерное на $(0, \frac{1}{2})$ и равномерное на $(\frac{1}{2}, 1)$ с вероятностью β_2 ($\beta_1 + \beta_2 = 1$). Здесь β_1 и β_2 можно воспринимать, например, как степень влияния положительной (или отрицательной) рекламы соответственно. Можно показать, что в данной модели профиль мнений $\vec{x}(t)$ является марковским процессом с дискретным временем и непрерывным пространством состояний.

Для изучения поведения профиля на больших временах введем аналог плотности мнений. А именно, обозначим через $\rho_k(t)$ - число агентов, мнения которых попали в интервал $\Delta_k = (\frac{k-1}{N}, \frac{k}{N})$, $k = \overline{1, N}$, $\rho_k(t) = \sum_{j=1}^n 1_{\{x_j(t) \in \Delta_k\}}$, $k = 1, \dots, N$. Заметим, что случайный процесс (с.п.) $\rho_k(t)$ уже не является марковским. На рис.1. представлены траектории конфигураций при различных значениях уровня доверия и графики плотности мнений для этих конфигураций в конкретные моменты времени t .

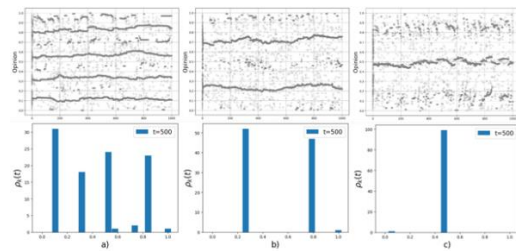


Рис.1. Эволюция конфигурации и эволюция плотности мнений на больших временах. $n = 100$ агентов, $\mu = 0.01$, $\beta_1 = 0.5$, $t = 500$. а) $\varepsilon = 0.1$, б) $\varepsilon = 0.2$, в) $\varepsilon = 0.3$

Мы наблюдаем уменьшение числа кластеров с ростом радиуса доверия, как и в классической модели Х-К с детерминированной динамикой, а наличие случайности влияет на существование малых кластеров. В данной модели мы также можем заметить явление, отмеченное в работе [10].

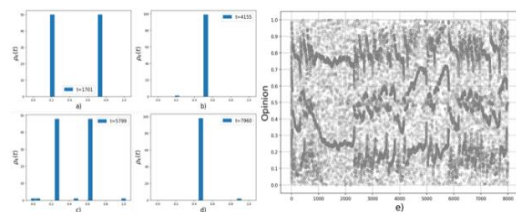


Рис.2. Эффект смены числа “кластеров”. $n = 100$ агентов. $\varepsilon = 0.242$, $\mu = 0.01$, $\beta_1 = 0.5$. а) $t = 1701$, б) $t = 4155$, в) $t = 5799$, д) $t = 7960$

При некоторых значениях радиуса доверия (рис.2) (например, $\varepsilon = 0.242$) кластеры перестают быть стабильными, и мы наблюдаем периодическую смену числа кластеров с одного на два и наоборот. Для более точного наблюдения за сменой числа кластеров на траектории системы определим с.п., который будем называть параметром порядка на траектории с.п. $\bar{x}(t)$ [11]:

$$Q(t) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n 1_{\{|x_i(t) - x_j(t)| \leq \varepsilon\}}$$

$Q(t)$ характеризует размер наибольшего кластера в группе в момент времени t . Если в момент t один кластер, то $Q(t) = 1$, если два равных кластера, то $\frac{1}{2}$.

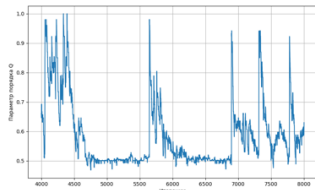


Рис.3. График зависимости параметра порядка Q от числа итераций.
 $n = 100$ агентов, $\varepsilon = 0.242$, $\mu = 0.01$

На рис.3 мы наблюдаем пики траектории $Q(t)$, которые соответствуют тому, что все агенты находятся в области взаимодействия. Также присутствуют большие участки траектории на уровне 0,5, что соответствует двум почти равным кластерам.

Модель 2 с ограниченным доверием с аддитивным шумом

В данной модели, в отличие от модели 2, внешнее влияние учитывается в виде аддитивных слагаемых на каждом шаге алгоритма. То есть, каждый агент (независимо от других) подвержен случайному влиянию на каждом шаге. Алгоритм имеет следующий вид:

$$x_i(t+1) = \frac{\alpha^0}{N_i(t)} \sum_{j \in N_i(t)} x_j(t) + \alpha^1 \xi^1(t) + \alpha^2 \xi^2(t),$$

где $\alpha^j \geq 0$, $\alpha^0 + \alpha^1 + \alpha^2 = 1$, $\xi_i^k(t)$ – н.с.в., $\xi_i^1(t) \sim U(0, \frac{1}{2})$, $\xi_i^2(t) \sim U(\frac{1}{2}, 1)$.

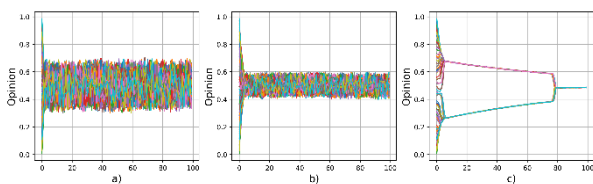


Рис.4. Траектории модели 2. $n = 100$, $\varepsilon = 0.2$
a) $\alpha^0 = 0.2, \alpha^1 = 0.4, \alpha^2 = 0.4$,
b) $\alpha^0 = 0.6, \alpha^1 = 0.2, \alpha^2 = 0.2$,
c) $\alpha^0 = 0.99, \alpha^1 = 0.005, \alpha^2 = 0.005$

Результаты моделирования (рис.4) демонстрируют, что в данной модели, в отличие от модели 1, всегда образуется один квазикластер, однако размах кластера разный. Для его качественного анализа определим с.п. – диаметр конфигурации в момент времени t : $d(t) = \max_i(x_i(t)) - \min_i(x_i(t))$. Мы будем изучать временное среднее диаметра $d_T = \frac{1}{T} \sum_{t=0}^T (\max_i(x_i(t)) - \min_i(x_i(t)))$. Было выявлено, что значение d_T стабилизируется на больших временах.

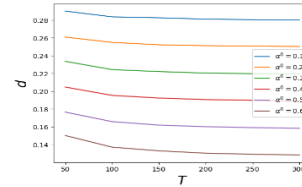


Рис.5. Зависимость диаметра квазикластера от T .
 $n=100$, $\varepsilon = 0.2$, $\alpha^1 = 0.5\alpha$, $\alpha^2 = 0.5\alpha$

Величина диаметра d_T зависит от параметра α^0 , характеризующего взаимодействие согласно Х-К. На рис.6 приведены результаты моделирования для $\alpha^0 \in (0,1)$ и трех разных соотношениях α^1 и α^2 ($\alpha^1 + \alpha^2 = \alpha = 1 - \alpha^0$). Заметим, что d_T увеличивается с уменьшением α^0 , что может иметь вполне естественную интерпретацию: чем меньше α^0 , тем больше случайное влияние на систему, и агенты больше подвержены резкому изменению своего мнения. Кроме того, заметим, что при фиксированном α^0 диаметр d_T минимален при $\alpha^1 = \alpha^2$, т. к. при этом дисперсия случайного воздействия минимальна.

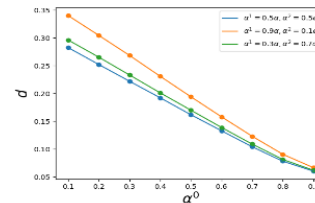


Рис.6. Зависимость диаметра притягивающего множества от параметра шума.
 $n = 100$, $\varepsilon = 0.2$, $T = 100$

Проверим гипотезу об эргодичности рассматриваемой системы. Для этого рассмотрим различные распределения начального профиля. Для этого используем класс бета-распределений на отрезке $[0,1]$ с различными параметрами. Зафиксируем $\varepsilon = 0.2$, $\alpha^1 = 0.2$, $\alpha^2 = 0.4$. На рис.7 приведены результаты моделирования для начальных распределений: $x(0) \sim B(1,100)$, $B(100,1)$, $B(1,1)$, $U(0,1)$. Можно наблюдать, что на больших временах поведение профиля не зависит от начального распределения, что является проявлением эргодичности системы.

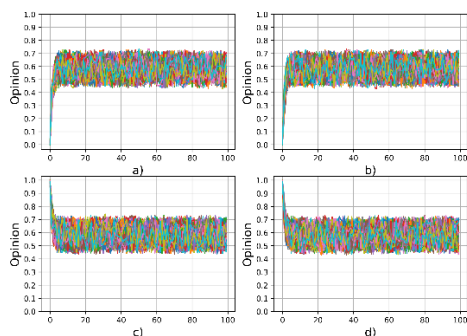


Рис.7. Эволюция конфигураций для модели с аддитивным шумом.

$n = 100, \varepsilon = 0.2, \alpha^1 = 0.2, \alpha^2 = 0.4,$
 a) $x(0) \sim B(1, 100)$, b) $x(0) \sim B(100, 1)$,
 c) $x(0) \sim B(1, 1)$, d) $x(0) \sim U(0, 1)$

Заключение

В данной работе были сформулированы новые стохастические модели динамики мнений, основанные на алгоритме Хегсельмана-Краузе, которые позволяют учесть влияние различных источников информации. Результаты моделирования показали, что если агенты могут случайно менять свое мнение, то качественно поведение системы похоже на поведение классической модели Х-К. Однако наблюдается эффект смены числа кластеров при некоторых значениях уровня доверия. Для модели с постоянным влиянием нескольких внешних источников образуется один квазикластер, диаметр которого зависит как от распределения внешних воздействий, так и от восприимчивости агентов к таким воздействиям. Данные модели, по нашему мнению, с одной стороны, позволяют учитывать случайную природу процесса взаимодействия и принятия решений, с другой стороны, могут быть использованы при исследовании открытых многоагентных систем.

Список литературы

1. Zha Q., Kou G., Zhang H., Liang H., Chen X., Li C., Dong Y. Opinion dynamics in finance and business // *Financial Innovation*. – 2020. – V. 6. – № 44.
2. Peralta A.F., Kertesz J., Iniguez G. Opinion dynamics in social networks: From models to data // *arXiv:2201.01322v4*. – 2022
3. Proskurnikov A.V, Tempo R. A Tutorial on Modeling and Analysis of Dynamic Social Networks. Part II // *Annual Reviews in Control*. – 2018. – V. 45. – P.166-190.
4. Kerckhove C.V., Martin S., Gend P., Rentfrow P.J., Hendrickx J.M., Blondel V.D. Modelling influence and opinion evolution in online collective behaviour // *PLoS One*. – 2016. – V. 11. – № 6. – P. 1-25.
5. Quattrocchi W., Scala A., Caldarelli G. Opinion dynamics on interacting networks: Media competition and social influence // *Scientific Reports*. – 2014. – V. 4. – P. 1-7.
6. Yang B., Yu Q., Fan Y. A Hybrid Opinion Formation and Polarization Model // *Entropy*. – 2022. – V. 24. – № 11.

7. Hegselmann R., Krause U. Opinion dynamics and bounded confidence models, analysis, and simulation // *JASSS*. – 2002. – V. 5. – №. 3. – P. 1-33.

8. Su W., Chen G., Hong Y. Noise leads to quasi-consensus of H-Krause opinion dynamics // *Automatica*. – 2017. – V. 85. – P. 448-454.

9. Nyczka P. A Model of Opinion Dynamics with Bounded Confidence and Noise // <https://arxiv.org/abs/1106.0008>

10. Pineda M., Toral R., Hernandez-Garcia E. The noisy H-Krause model for opinion dynamics // *The European Physical Journal B*. – 2013. – V. 46. – № 490.

11. C. Wang, Q. Li, W. E, B. Chazelle. Noisy H-Krause systems: Phase transition and the 2 R-conjecture. // *Journal of Statistical Physics*. – 2017. – V. 166. – P. 1209 -1225.

ВЫВОД УРАВНЕНИЙ ДВУХПАЛУБНОГО ОБТЕКАНИЯ НЕРОВНОЙ ПОВЕРХНОСТИ С МАЛЫМИ БЫСТРО ОСЦИЛЛИРУЮЩИМИ НЕРОВНОСТЯМИ

Глазунова А.М.

Национальный исследовательский университет
 «Высшая школа экономики»,
 департамент прикладной математики
 МИЭМ НИУ ВШЭ

Аннотация

В работе изучается обтекание вязкими несжимаемыми жидкостями криволинейных поверхностей, геометрия которых приводит к возникновению в потоке двухпалубной структуры, при больших значениях числа Рейнольдса. Проводится сравнительный анализ с подобной задачей обтекания пластины.

Введение

Основным предметом исследования является пограничный слой — тонкий слой около поверхности, обтекаемой вязкой жидкостью, в котором скорость течения возрастает от нулевого значения до значения скорости внешнего потока [1]. Течение в пограничном слое влияет на величину сопротивления трения, что приводит к необходимости определения поля скоростей в нем для различных задач. Выбор формы поверхности для исследования обусловлен тем, что в прикладных задачах поверхность чаще не плоская, а изогнутая, а также может содержать небольшие неровности, например, заклёпки на корпусе или шероховатости, возникающие из-за нанесённых слоёв краски [2]. Более того, для наиболее общей модели описания задач гидродинамики — уравнений Навье-Стокса — не доказаны существование и единственность решения в неограниченных областях, поэтому решение поставленной задачи требует применение различных аналитических и численных методов. Таким образом, целью исследования является описание двухпалубного обтекания криволинейной поверхности с малыми возмущениями при больших значениях числа Рейнольдса.

Системы уравнений для различных областей двухпалубной структуры

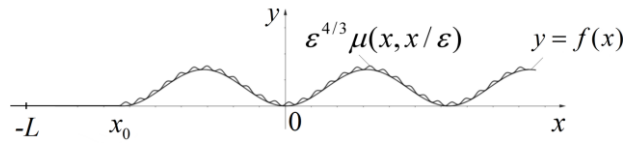


Рис.1. Исследуемая поверхность

На поверхность, представляющую собой часть полуплоскости $y = 0$, $x < x_0$ гладко продолженную кривой $y = f(x)$, набегает параллельный внешний поток. Поверхность содержит малые неровности, которые задаются функцией $\varepsilon^{4/3} \mu(x, x/\varepsilon)$, $\varepsilon = 1/\sqrt{Re}$, где Re — число Рейнольдса, а функция $\mu(x, \xi) \in C^\infty([x_0, \infty] \times [0, 2\pi])$, $\xi = x/\varepsilon$ обладает свойствами:

$$\int_0^{2\pi} \mu(x, \xi) d\xi = 0,$$

$$\left. \frac{\partial^k \mu}{\partial x^k}(x, \xi) \right|_{x=x_0} = 0, k \leq 2.$$

Параметры неровности выбраны таким образом, что обтекающий поток при $f(x) = 0$ имеет двухпалубную структуру.

Таким образом, общая формула, задающая поверхность обтекания, имеет вид:

$$y_s = f(x) + \varepsilon^{4/3} \mu(x, x/\varepsilon).$$

Для рассматриваемой задачи мы получили системы уравнений, описывающие течение во всех слоях двухпалубной структуры, а именно в области внешнего течения, области классического и тонкого пограничных слоёв (принадлежность функции к определённой области (палубе) выражено верхним индексом: III, II или I соответственно) [3].

Отметим, что все выведенные уравнения в предельном случае обтекания поверхности нулевой кривизны сходятся с уравнениями, полученными для задачи обтекания пластины с малыми периодическими неровностями [4]. Наличие дополнительных слагаемых в полученных уравнениях приводит к тому, что системы уравнений, описывающие течение в различных областях, оказываются связаны между собой.

Выражения для скоростей во внешнем потоке

В отличие от задачи обтекания пластины, для рассматриваемой задачи нахождение главных слагаемых асимптотического разложения требует решения системы, состоящей из уравнений Эйлера и уравнения неразрывности, дополненных граничным условием (1). Для задачи обтекания криволинейной поверхности (без малых неровностей) также была получена такая система, но не рассмотрено её решение [5].

$$\begin{cases} u_0 \frac{\partial u_0}{\partial x} + v_0 \frac{\partial u_0}{\partial y} = -\frac{\partial p_0}{\partial x}, \\ u_0 \frac{\partial v_0}{\partial x} + v_0 \frac{\partial v_0}{\partial y} = -\frac{\partial p_0}{\partial y}, \\ \frac{\partial u_0}{\partial x} + \frac{\partial v_0}{\partial y} = 0, \end{cases} \quad (1)$$

$$-y'_x u_0|_{y=f(x)} + v_0|_{y=f(x)} = 0.$$

Аналитические решения этой системы в общем случае не известны, поэтому возможны несколько способов дальнейшего решения. Первый способ — это применение численных методов решения, второй — рассматривать задачу, в которой поверхность $y = f(x)$ мало отличается от плоской, $y = \delta f(x)$, где δ — малый параметр, не связанный с числом Рейнольдса. Тогда в главном, в соответствии с рис. 1, решение системы (1) будет иметь вид

$$\begin{pmatrix} u'_0 \\ v'_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

В таком случае выражения для скоростей можно найти аналитически, используя асимптотическое разложение по малому параметру δ :

$$\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} = \begin{pmatrix} 1 + \delta u_1 + \delta^2 u_2 + O(\delta^3) \\ \delta v_1 + \delta^2 v_2 + O(\delta^3) \end{pmatrix},$$

(ограничимся только первой поправкой). Для нахождения первых поправок введём функцию тока Ψ :

$$u_1 = \frac{\partial \Psi}{\partial y}, v_1 = -\frac{\partial \Psi}{\partial x},$$

Граничное условие примет вид:

$$-\delta f'_x u_0|_{y=f(x)} + v_0|_{y=f(x)} = 0.$$

Значения главных слагаемых на границе равны 1 и 0 соответственно, поэтому при δ^1 остаётся равенство $v_1|_{y=f(x)} = f'_x$.

Исходя из этого, при степени δ , равной 1, система (1) приняла вид:

$$\begin{cases} \Delta \Psi = 0, & y > 0 \\ \Psi|_{y=f(x)} = -f(x) \end{cases}$$

Решение полученной системы с применением преобразования Фурье приводит к нахождению главных поправок u_1, v_1 (в предположении, что Ψ абсолютно интегрируема):

$$\begin{aligned} u_1 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |p| e^{-|p|y} e^{ip(x-t)} f(t) dt dp, \\ v_1 &= \frac{i}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p e^{-|p|y} e^{ip(x-t)} f(t) dt dp. \end{aligned} \quad (2)$$

Можно упростить полученные выражения (2), взяв интеграл по переменной \mathbf{p} :

$$\begin{aligned} u_1 &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{-(x-t)^2 + y^2}{((x-t)^2 + y^2)^2} f(t) dt, \\ v_1 &= -\frac{2y}{\pi} \int_{-\infty}^{+\infty} \frac{(x-t)f(t) dt}{((x-t)^2 + y^2)^2}. \end{aligned} \quad (3)$$

Используя эти формулы, можно найти решение системы уравнений из пограничного слоя.

Уравнения для скоростей в пограничных слоях

В пограничных слоях введены следующие быстрые переменные:

$$\tau = \frac{y - f(x)}{\varepsilon}, \quad \theta = \frac{y - f(x)}{\varepsilon^{4/3}}.$$

В классическом пограничном слое возникает задача решения уравнений, сводящихся в предельном случае (при $f'_x = 0$) к уравнениям Прандтля

$$\begin{cases} \left(\bar{u}_0^{II} + \bar{u}_0^{III} \Big|_{y=f(x)} \right) \frac{\partial \bar{u}_0^{II}}{\partial x} + \bar{u}_0^{II} \frac{\partial \bar{u}_0^{III}}{\partial x} \Big|_{y=f(x)} + \\ + \frac{\partial \bar{u}_0^{II}}{\partial \tau} \left(\sqrt{f'^2_x + 1} \bar{v}_2^{II} - f'_x \bar{u}_1^{III} \Big|_{y=f(x)} + \bar{v}_1^{III} \Big|_{y=f(x)} \right) = \\ = (f'^2_x + 1) \frac{\partial^2 \bar{u}_0^{II}}{\partial \tau^2}, \\ \frac{\partial \bar{u}_0^{II}}{\partial x} + \sqrt{f'^2_x + 1} \frac{\partial \bar{v}_2^{II}}{\partial \tau} = 0, \end{cases} \quad (4)$$

и уравнению типа Рэлея

$$\begin{aligned} \left(\bar{u}_0^{II} \Big|_{\tau=0} + \bar{u}_0^{III} \Big|_{y=f(x)} \right) \left(\frac{\partial^2 \bar{v}_1^{II}}{\partial \xi^2} - 2f'_x \frac{\partial^2 \bar{v}_1^{II}}{\partial \xi \partial \tau} \right. \\ \left. + (f'^2_x + 1) \frac{\partial^2 \bar{v}_1^{II}}{\partial \tau^2} \right) \\ - \bar{v}_1^{II} (f'^2_x + 1) \frac{\partial^2 \bar{u}_0^{II}}{\partial \tau^2} = 0; \end{aligned} \quad (5)$$

в тонком слое — системы уравнений Прандтля с самоиндуцированным давлением:

$$\begin{cases} u_1^* \frac{\partial u_1^*}{\partial \xi} + (f'^2_x + 1) v_1^* \frac{\partial u_1^*}{\partial \theta} - (f'^2_x + 1)^{3/2} \frac{\partial^2 u_1^*}{\partial \theta^2} \\ + (f'^2_x + 1) \left(\frac{\partial \bar{p}_2^{II}}{\partial \xi} \Big|_{\tau=0} - f'_x \frac{\partial \bar{p}_2^{II}}{\partial \tau} \Big|_{\tau=0} \right) = 0, \\ \frac{\partial u_1^*}{\partial \xi} + (f'^2_x + 1) \frac{\partial v_1^*}{\partial \theta} = 0, \end{cases} \quad (6)$$

где

$$\begin{aligned} u_1^* &= u_1^I + \bar{u}_1^{II} \Big|_{\tau=0} + \theta \sqrt{f'^2_x + 1} \frac{\partial \bar{u}_0^{II}}{\partial \tau} \Big|_{\tau=0}, \\ v_1^* &= \bar{v}_1^I + \bar{v}_1^{II} \Big|_{\tau=0}. \end{aligned}$$

Давление определяется из следующего равенства:

$$\begin{aligned} \frac{\partial \bar{p}_2^{II}}{\partial \xi} &= \\ &= \sqrt{f'^2_x + 1} \left[\bar{u}_0^{II} \Big|_{\tau=0} \right. \\ &+ \bar{u}_0^{III} \Big|_{y=f(x)} \left((f'^2_x + 1) \frac{\partial \bar{v}_1^{II}}{\partial \tau} \right. \\ &\left. \left. - f'_x \frac{\partial \bar{v}_1^{II}}{\partial \xi} \right) - \bar{v}_1^{II} (f'^2_x + 1) \frac{\partial \bar{u}_0^{II}}{\partial \tau} \right]. \end{aligned}$$

В уравнении Прандтля (4) можно выразить сумму $-f'_x \bar{u}_1^{III} \Big|_{y=f(x)} + \bar{v}_1^{III} \Big|_{y=f(x)}$, используя следующее граничное условие:

$$\begin{aligned} -f'_x \bar{u}_1^{III} \Big|_{y=f(x)} + \bar{v}_1^{III} \Big|_{y=f(x)} &= \\ = -\sqrt{f'^2_x + 1} \left(\bar{v}_2^{II} \Big|_{\tau=0} + \mu \frac{\partial \bar{v}_1^{II}}{\partial \tau} \Big|_{\tau=0} \right). \end{aligned}$$

Таким образом, получаем, что сцеплены 3 системы: уравнения Прандтля (4) и уравнение типа Рэлея (5) через функцию \bar{u}_0^{II} и среднюю часть произведения

$$\mu \frac{\partial \bar{v}_1^{II}}{\partial \tau} \Big|_{\tau=0};$$

а также система уравнений из тонкого слоя (6) и уравнение типа Рэлея (5) через граничные условия:

$$\frac{\partial u_1^*}{\partial \theta} \Big|_{\theta \rightarrow \infty} \rightarrow \sqrt{f'^2_x + 1} \frac{\partial \bar{u}_0^{II}}{\partial \tau} \Big|_{\tau=0}, \quad v_1^* \Big|_{\theta \rightarrow \infty} \rightarrow \bar{v}_1^{II} \Big|_{\tau=0}.$$

Можно расцепить эти уравнения итерационным методом. Также, если рассматривать поверхности, незначительно отличающиеся от плоскости, можно ограничиться приближенным решением уравнения Эйлера (главным слагаемым асимптотического разложения и первой поправкой (3)).

Граничные условия

Полученные ранее уравнения дополняются граничными условиями. Приведём полученные равенства:

$$\begin{cases} \bar{u}_0^{II} \Big|_{\tau=0} + \bar{u}_0^{III} \Big|_{y=f(x)} = 0, \\ \bar{v}_0^{II} \Big|_{\tau=0} + \bar{v}_0^{III} \Big|_{y=f(x)} = 0. \end{cases}$$

$$\left(u_1^I \Big|_{\theta=\mu} + \bar{u}_1^{II} \Big|_{\tau=0} \right) \frac{1}{\sqrt{f'^2_x + 1}} + \mu \frac{\partial \bar{u}_0^{II}}{\partial \tau} \Big|_{\tau=0} = 0.$$

Из последнего равенства выводится граничное условие для функции u_1^* :

$$u_1^* \Big|_{\theta=\mu} = u_1^I \Big|_{\theta=\mu} + \bar{u}_1^{II} \Big|_{\tau=0} + \theta \sqrt{f'^2_x + 1} \frac{\partial \bar{u}_0^{II}}{\partial \tau} \Big|_{\tau=0} = 0.$$

Равенство

$$\tilde{v}_1^I|_{\theta=\mu} + \tilde{v}_1^{II}|_{\tau=0} = 0$$

позволяет найти граничное условие для функции v_1^* :

$$v_1^*|_{\theta=\mu} = \tilde{v}_1^I|_{\theta=\mu} + \tilde{v}_1^{II}|_{\tau=0} = 0.$$

$$\left\{ \begin{array}{l} \bar{u}_1^{III}|_{y=f(x)} + f'_x \bar{v}_1^{III}|_{y=y(x)} + \sqrt{f'^2_x + 1} \left(u_3^I|_{\theta=\mu} + \right. \\ \left. + u_3^{II}|_{\tau=0} + \frac{1}{2} \mu^2 \frac{\partial^2 \bar{u}_1^{III}}{\partial \tau^2} \Big|_{\tau=0} + \mu \frac{\partial u_2^{II}}{\partial \tau} \Big|_{\tau=0} \right) + \\ \left. + \frac{1}{6} \mu^3 (y'^2_x + 1) \frac{\partial^3 \bar{u}_0^{III}}{\partial \tau^3} \Big|_{\tau=0} \right. = 0, \\ -f'_x \bar{u}_1^{III}|_{y=f(x)} + \bar{v}_1^{III}|_{y=f(x)} + \\ \left. + \sqrt{f'^2_x + 1} \left(\tilde{v}_2^I|_{\theta=\mu} + \bar{v}_2^{II}|_{\tau=0} + \mu \frac{\partial \tilde{v}_1^{II}}{\partial \tau} \Big|_{\tau=0} \right) \right. = 0. \end{array} \right.$$

Заключение

В работе приведены системы уравнений, возникающие при асимптотической редукции полной системы уравнений Навье–Стокса в задаче о течении несжимаемой вязкой жидкости при больших значениях числа Рейнольдса вдоль криволинейной поверхности с малыми периодическими неровностями. Этот результат является обобщением работы (4), в которой была описана двухпалубная структура пограничного слоя, возникающая при обтекании пластины.

Работа основана на совместных результатах с В.Г. Даниловым.

Список литературы

1. Prandtl, L. Uber Flfissigkeitsbewegung bei sehr kleiner Reibung / L. Prandtl // 3rd International Congress of Mathematicians, 1904. - P. 484-491.
2. Шлихтинг, Г. Теория пограничного слоя / Г. Шлихтинг. - пер. Г. А. Вольперта с 5-го нем. изд., испр. по 6-му (амер.) изд. - Москва: Наука, 1974. - С. 589.
3. Danilov, V. G., Glazunova, A. M. Flow Around a Curved Plate with Small Periodic Irregularities: a Double-Deck Boundary Layer / V. G. Danilov, A. M. Glazunova // Russian J. of Mathematical Physics. - 2023. - Vol. 30. - № 4. P. 453-465.
4. Danilov, V. G. Asymptotic and numerical analysis of the flow around a plate with small periodic irregularities / V. G. Danilov, M. V. Makarova // Russian J. of Mathematical Physics. - 1994. - Vol. 2. - № 1. - P. 49-56.
5. Maslov, V. P., Omel'yanov G. A. Lin-Lees equations for boundary layers in domains with curvilinear boundary / V. P. Maslov, G. A. Omel'yanov // Physica D: Nonlinear Phenomena. - 1988. - Vol. 33. - P. 266-280.

ОБ ОБОСНОВАНИИ НЕОСЦИЛЛИРУЮЩИХ ВКБ АСИМПТОТИК

Петухова Е.С.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ

Аннотация

В данной работе разрабатываются методы обоснования асимптотических неосциллирующих ВКБ решений для задачи о случайном блуждании по одномерной решетке. Разностное уравнение для решетки продолжается на всю прямую (конфигурационное пространство), для полученного соответствующего псевдодифференциального уравнения строится решение задачи Коши с помощью свертки с фундаментальным решением. В этом случае, оценка разности между точным и асимптотическим ВКБ решениями задачи находится довольно легко - в смысле нормы максимума модуля, но получение оценки в норме максимума модуля с весом равным обратной экспоненте - не такая простая задача.

Введение

В работе рассматривается задача Коши для разностного уравнения, описывающего случайное блуждание по решетке:

$$\begin{aligned} U_i^{n+1} &= a_{i,i-k} U_{i-k}^n + \dots + a_{i,i+m} U_{i+m}^n \\ U_i^0 &= g_i \end{aligned} \quad (1)$$

где $a_{i,j}$ - вероятности перехода из состояния j в состояние i , удовлетворяющие условиям $a_{i,j} \geq 0$, $\sum_i a_{i,j} = 1$.

Решение задачи характеризует вероятность попадания блуждающей частицы в некоторую точку пространства в определенный момент времени.

Хорошо известно, что асимптотическое решение задачи (1) допускает оценку

$$\begin{aligned} \max_i |U_i^n - \bar{U}_i^n| &\leq \\ \max_i |\bar{U}_i^0| + \sum_{k=0}^n \max_i |F_i^k| & \end{aligned} \quad (2)$$

где F_i^n - некоторая невязка неоднородного уравнения, возникающая вследствие приближенности решения.

Если $F_i^n = h^2 * R_i^n$ (такая правая часть возникает при построении асимптотического решения $\bar{U}(x, t)$), то для разности точного и асимптотического решений получаем оценку $\max_i |U_i^n - \bar{U}_i^n| \leq h^2 \sum_{i=0}^n R_i^n$. Так как количество слагаемых в сумме $n \leq \frac{t}{h}$, где h - шаг сетки, то окончательно получаем $\max_i |U_i^n - \bar{U}_i^n| \leq O(h)$.

К сожалению, оценка, приведенная выше, не дает информацию о точности решения в области $\bar{U}(x, t) > O(h^{1+\gamma})$, так как отброшенные в ходе построения

слагаемые оказываются формально больше, чем асимптотическое решение.

В работе это противоречие устраняется с помощью более точной оценки невязки.

Постановка задачи

Простейшее случайное блуждание по решетке, с конечным числом шагов, описывается разностным уравнением (1).

Совершим переход к псевдодифференциальным уравнениям для данной задачи. Для этого заменим функцию дискретного аргумента U_i^n функцией непрерывного аргумента $U(x, t)$, разностные уравнения – псевдодифференциальными.

Для функции непрерывного аргумента должно выполняться равенство $U_i^n = U(ih, n\tau)$, где τ – длина шага по времени. Тогда уравнение, описывающее блуждание будет иметь вид:

$$e^{\tau \frac{\partial}{\partial t}} U = \left(\sum_{j=-k}^m a_j(x) e^{jh \frac{\partial}{\partial x}} \right) U, \quad (3)$$

где $e^{kh \frac{\partial}{\partial x}}(f) = f(x + kh)$ – оператор сдвига.

Уравнение (3) переходит в уравнение (1), если рассматривать уравнение (3) в узлах сетки, то есть в точках $t_n = n\tau$, $x_i = ih$ и предполагать, что $a_j(ih) = a_{i,j}$, $U(ih, 0) = U_i^0$. При этом сужение на сетку решения уравнения (3) будет являться решением уравнения (1).

Решение задачи (3) в данной работе будет иметь вид неосциллирующего ВКБ решения:

$$U(x, t) = e^{-\frac{S(x,t)}{h}} \varphi(x, t).$$

Далее в работе для удобства вычислений сделано предположение $\tau = h$, то есть шаг по времени равен шагу сетки.

Построение асимптотического решения псевдодифференциального уравнения

В качестве первого шага построения решения уравнения (3) с начальным условием $U(x, 0)$ мы вводим понятие фундаментального решения задачи Коши $G(x, \xi, t)$, которое является решением следующей задачи:

$$\begin{aligned} G(x, \xi, t + h) &= \sum_{j=-k}^m e^{jh \frac{\partial}{\partial x}} (a_j(x) G(x, \xi, t)) = \\ &= \sum_{j=-k}^m a_j(x + h) G(x + h, \xi, t), \\ G(x, \xi, 0) &= \delta(x - \xi) \end{aligned}$$

Формула для фундаментального решения псевдодифференциального уравнения, которая будет использоваться, имеет вид:

$$G(x, \xi, t + h) = \lim_{\beta \rightarrow 1-0} G_\beta(x, \xi, t),$$

где $G_\beta(x, \xi, t) = e^{-\frac{\Phi_\beta(x, \xi + y, t) + \frac{y^2}{2}}{h}} \widehat{\varphi}_\beta(x, \xi, t) (1 + O(h))$,

$y = \frac{\partial \Phi_\beta}{\partial y}(x, \xi + y, t)$ и Φ_β – решение уравнения Гамильтона-Якоби, отвечающего уравнению (3) с начальным условием

$$\Phi_\beta|_{t=0} = \frac{\beta}{2} (x - y - \xi)^2$$

Когда $\beta < 1$, функция $G_\beta(x, \xi, t)$ гладкая и $LG_\beta = h^2 R_\beta$, где L – оператор, входящий в уравнение (3), а R_β – ограниченная при $\beta < 1$ функция. В случае, когда $a_j(x) = const$ оператор L имеет вид:

$$L = e^{h \frac{\partial}{\partial t}} - \left(\sum_{j=-k}^m a_j e^{jh \frac{\partial}{\partial x}} \right)$$

Так как $G_\beta(x, \xi, t)$ гладкая функция, то обозначая

$$S_\beta = \Phi_\beta(x, \xi + y, t) - \frac{y^2}{2}, y = \frac{\partial \Phi_\beta}{\partial y}(x, \xi + y, t)$$

по теореме об оценке разности между точным и асимптотическим решением получим, что

$$\max_i |W_i^{n+1}| \leq \frac{1}{2} \int_0^t \max_x (S_{\beta t t} - S_{\beta x x}) dt' (h * \max_\Omega |R_\beta|),$$

где

$$W = \left| e^{\frac{S_\beta}{h}} (\overline{G}_\beta - G_\beta) \right| \text{ и } \overline{G}_\beta - \text{точное решение, } \Omega = \{x \in R, h \leq t \leq nh\}.$$

Пусть $U(x, 0) = e^{-\frac{S_0}{h}} \varphi_0$, тогда асимптотическое решение задачи Коши для псевдодифференциального уравнения имеет вид:

$$U_{as}(x, t) = \lim_{\beta \rightarrow 1-0} \int_{\mathbb{R}^1} G_\beta(x - \xi, t) U(\xi, 0) d\xi.$$

В этом случае можно перейти к пределу под знаком интеграла при $t > 0$ и, с помощью приведенной выше оценки, получить оценку решения задачи Коши:

$$\max_i \left| (U_{as} - U_{\text{точн.}}) e^{\frac{S}{h}} \right| \leq e^{\frac{1}{2} \int_0^t \max_x (S_{tt} - S_{xx}) dt'} O(h) \quad (4)$$

где $S = S(x, t)$ – решение задачи Коши для уравнения Гамильтона-Якоби с начальным условием $S_0(x)$. Если же $U(x, 0) = \frac{1}{h} \Pi(x)$, где

$$\Pi(x) = \begin{cases} 1, & |x| < \frac{h}{2} \\ 0, & |x| > \frac{h}{2}, \\ \frac{1}{2}, & |x| = \frac{h}{2} \end{cases}$$

то функция $G_\beta(x - \xi, t)$ в пределе при $\beta = 1 - 0$ имеет интегрируемую особенность типа $1/\sqrt{\quad}$ на границе области достижимости случайного блужда-

ния. Это приводит к тому, что в правой части оценки (4) появляется $\sqrt{h} \ln h$ вместо $O(h)$.

Заключение

Таким образом, в работе описаны метод построения асимптотического решения задачи Коши для уравнений, описывающих случайное блуждание на решетке, и способ оценки точности данного ВКБ решения. Показано, что построенное неосциллирующим методом ВКБ асимптотическое решение приближает точное с экспоненциальной точностью.

Доклад основан на результатах, полученных совместно с В.Г. Даниловым.

Список литературы

1. Danilov V. G. A representation of the delta-function via creation operators and gaussian exponentials, and multiplicative fundamental solution asymptotics for some parabolic pseudodifferential-equations //Russian Journal of Mathematical Physics. – 1995. – Т. 3. – №. 1. – С. 25-40.
2. Danilov V. G. Nonsmooth nonoscillating exponential-type asymptotics for linear parabolic PDE //SIAM Journal on Mathematical Analysis. – 2017. – Т. 49. – №. 5. – С. 3550-3572.
3. Маслов В. П. Асимптотические методы и теория возмущений. – «Наука», Главная редакция физико-математической литературы, 1988.
4. Михайлова С.О., Петухова Е.С. Асимптотика решений уравнений на сетках в дискретном времени //Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов имени Е.В. Арменского. Материалы конференции. М.: МИЭМ НИУ ВШЭ, 2023.

ПРИМЕНЕНИЕ ПОЛУМАРКОВСКИХ ПРОЦЕССОВ ДЛЯ МОДЕЛИРОВАНИЯ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ G|G|n

Шумов Н.А.

Национальный исследовательский университет «Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ

Аннотация

В представленной работе рассматривается система массового обслуживания (далее СМО) G|G|n, производится построение её математической модели и получение формул для ряда характеристик указанной СМО. Адекватность представленных формул подтверждается имитационным моделированием.

Введение

Исследование характеристик систем массового обслуживания производится с помощью имитационного моделирования [1] и аналитически [2-5] с целью решения прикладных задач. В настоящее время по-

лучение характеристик СМО G|G|n производится исключительно с помощью имитационного моделирования, что может влиять на точность полученных результатов.

Разработанная математическая модель позволяет получить ряд характеристик СМО в аналитическом виде.

В данной работе также осуществляется вывод формул для нахождения стационарного и предельного распределений и получение условий их существования.

Математическая модель

В представленной работе построена математическая модель СМО G|G|n в виде двумерного полумарковского процесса $\xi(t) \in E$, E - множество пар (i, r) где i - число требований в системе в момент времени t , $r = 0$, если переход в i произошел после окончания обслуживания, $r = 1$, если переход в i произошел после прихода требования. Моменты t_n прихода и ухода требований для указанного двумерного процесса являются марковскими. Обозначим $\xi_n = (\xi(t_n + 0), r(t_n + 0))$ - вложенная марковская цепь.

Введем следующие обозначения: пусть $F(t)$ - функция распределения интервалов между поступлениями заявок, $G(t)$ - функция распределения времени обслуживания одной заявки в любом канале обслуживания, $\eta_X(t)$ - функция распределения времени перескока,

$B_i(x) = \left[\frac{\int_0^x \bar{G}(t) dt}{\int_0^\infty \bar{G}(t) dt} \right]^i$, $B_n(x) = \left[\frac{\int_0^x \bar{G}(t) dt}{\int_0^\infty \bar{G}(t) dt} \right]^n$, где $\bar{B}_i(x)$ и $\bar{B}_n(x)$ - функции распределения минимума из времен перескока при, соответственно, i или n занятых каналах.

Переходные вероятности вложенной марковской цепи $P_{ij}(ind) = P(\xi_{n+1} = j | \xi_n = i, r = ind)$, $ind \in \{0, 1\}$ имеют вид:

$$\begin{cases}
 P_{ij}(ind) = \\
 \left. \begin{aligned}
 &0, |j - i| \geq 2, \\
 &p_i^{(0,+)} = \int_0^\infty B_i(x) d\eta(x), j = i + 1, ind = 0, i < n \\
 &p_i^{(1,+)} = \int_0^\infty B_{i-1}(x) \bar{G}(x) dF(x), j = i + 1, ind = 1, i \leq n \\
 &p_{n+1}^{(0,+)} = \int_0^\infty B_{n-1}(x) \bar{G}(x) d\eta(x), j = i + 1, ind = 0, i \geq n \\
 &p_{n+1}^{(1,+)} = \int_0^\infty B_n(x) dF(x), j = i + 1, ind = 1, i > n \\
 &p_i^{(0,-)} = 1 - \int_0^\infty B_i(x) d\eta(x), j = i - 1, ind = 0, i < n \\
 &p_i^{(1,-)} = 1 - \int_0^\infty B_{i-1}(x) \bar{G}(x) dF(x), j = i - 1, ind = 1, i \leq n \\
 &p_{n+1}^{(0,-)} = 1 - \int_0^\infty B_{n-1}(x) \bar{G}(x) d\eta(x), j = i - 1, ind = 0, i \geq n \\
 &p_{n+1}^{(1,-)} = 1 - \int_0^\infty B_n(x) dF(x), j = i - 1, ind = 1, i > n
 \end{aligned}
 \right\} \quad (1)
 \end{cases}$$

Стационарное распределение

Обозначим $Q = ((q_0^{(0)}, q_0^{(1)}), (q_1^{(0)}, q_1^{(1)}), \dots)$ - стационарное распределение вложенной марковской цепи.

Утверждение 1.

Компоненты стационарного распределения вложенной марковской цепи являются геометрической прогрессией при $j \geq n + 1$, если $r = 0$, и при $j \geq n + 2$, если $r = 1$:

$$q_{j+1}^{(r)} = c_r q_j^{(r)}, r \in \{0,1\}, 0 < c_r < 1 \quad (2)$$

Следствие 1.

Система уравнений для нахождения стационарно-го распределения вложенной марковской цепи имеет вид:

$$\begin{cases} q_0^{(0)} = q_1^{(0)} p_1^{(0,-)} + q_1^{(1)} p_1^{(1,-)} \\ q_0^{(1)} = 0 \\ q_1^{(0)} = q_2^{(0)} p_2^{(0,-)} + q_2^{(1)} p_2^{(1,-)} \\ q_1^{(1)} = q_0^{(0)} \\ q_i^{(0)} = q_{i+1}^{(0)} p_{i+1}^{(0,-)} + q_{i+1}^{(1)} p_{i+1}^{(1,-)} \\ q_i^{(1)} = q_{i-1}^{(0)} p_{i-1}^{(0,+)} + q_{i-1}^{(1)} p_{i-1}^{(1,+)}, 2 \leq i \leq n \\ q_{n+1}^{(0)} = q_{n+1}^{(0)} c_0 p_{n+1}^{(0,-)} + q_{n+2}^{(1)} p_{n+1}^{(1,-)} \\ q_{n+1}^{(1)} = q_n^{(0)} p_n^{(0,+)} + q_n^{(1)} p_n^{(1,+)} \\ q_{n+2}^{(1)} = q_{n+1}^{(0)} p_{n+1}^{(0,+)} + q_{n+1}^{(1)} p_{n+1}^{(1,+)} \\ c_1 q_{n+2}^{(1)} = q_{n+1}^{(0)} c_0 p_{n+1}^{(0,+)} + q_{n+2}^{(1)} p_{n+1}^{(1,+)} \\ \frac{1}{1-c_0} q_{n+1}^{(0)} + \sum_{i=0}^n q_i^{(0)} + \frac{1}{1-c_1} q_{n+2}^{(1)} + \sum_{i=0}^{n+1} q_i^{(1)} = 1 \end{cases} \quad (3)$$

Утверждение 2.

Необходимое и достаточное условие существования стационарного распределения вложенной марковской цепи имеет вид:

$$0 < \frac{p_{n+1}^{(1,+)}}{p_{n+1}^{(0,-)}} < 1 \quad (4)$$

То есть в состоянии (n+1) вероятность прихода очередного требования должна быть меньше, чем вероятность окончания обслуживания в системе.

Предельное распределение

В прикладных задачах чаще требуется предельное распределение процесса $\xi(t)$, $\Pi = ((\pi_0^{(0)}, \pi_0^{(1)}), (\pi_1^{(0)}, \pi_1^{(1)}), \dots)$ компоненты которого имеют вид:

$$\pi_j^{(r)} = \frac{\int_0^\infty \bar{F}_j(x) dx q_j^{(r)}}{\sum_{i=0}^\infty \int_0^\infty \bar{F}_i(x) dx q_i^{(0)} + \sum_{i=0}^\infty \int_0^\infty \bar{F}_i(x) dx q_i^{(1)}}$$

где $r \in \{0,1\}$,

$\pi_j = \lim_{t \rightarrow \infty} P(\xi(t) = (j, 0)) + \lim_{t \rightarrow \infty} P(\xi(t) = (j, 1)) = \pi_j^{(0)} + \pi_j^{(1)}$ - предельное распределение первой компоненты, т.е. числа требований в системе, $F_i(x)$ – функция распределения времени пребывания в состоянии, зависит от последнего события в системе:

$$F_i(x, ind) = \begin{cases} 0, |j - i| \geq 2 \\ P(\min\{\eta_{Y_1}, \dots, \eta_{Y_k}, \eta_X\} < x) \\ P(\min\{\eta_{Y_1}, \dots, \eta_{Y_{k-1}}, Y_k, X\} < x) \\ P(\min\{\eta_{Y_1}, \dots, \eta_{Y_{n-1}}, Y_n, \eta_X\} < x) \\ P(\min\{\eta_{Y_1}, \dots, \eta_{Y_n}, X\} < x) \end{cases} = \quad (5)$$

$$= \begin{cases} 0, |j - i| \geq 2 \\ 1 - B_i(x) \bar{\eta}_X(x), j = i \pm 1, ind = 0, i < n \\ 1 - B_{i-1}(x) \bar{G}(x) \bar{F}(x), j = i \pm 1, ind = 1, i < n \\ 1 - B_{n-1}(x) \bar{G}(x) \bar{\eta}_X(x), j = i \pm 1, ind = 0, i \geq n \\ 1 - B_n(x) \bar{F}(x), j = i \pm 1, ind = 1, i \geq n \end{cases}$$

Заключение

В результате исследования получен аналитический вид предельного и стационарного распределений и условие их существования, смысл которого состоит в том, что в среднем заявки поступают реже, чем успевают обслуживаться.

Список литературы

1. Хмелевская А. В., Коптев Д. С., Шевцов А. Н., Щитов А. Н. Вариант алгоритма моделирования системы массового обслуживания общего вида (G/G/N/K) с произвольным законом распределения поступающих заявок // Известия Юго-Западного государственного университета, серия управление, вычислительная техника, информатика. Медицинское приборостроение. - 2016. - № 3.
2. Тарасов В.Н., Карташевский И.В., Липилина Л.В. Исследование задержки в системе G/G/1 // «Инфокоммуникационные технологии» - 2015. - 13 Т. - № 2. - с. 153-159.
3. Русилко Т. В., Сальников Д. А. Метод гауссова приближения для определения плотности вероятности вектора состояния сети массового обслуживания. – 2022.
4. Бочаров П.П., Печинкин А.В. Теория массового обслуживания: Учебник. - М.: Изд-во РУДН, 1995.
5. Блатов И. А., Карташевский В. Г., Киреева Н. В. Метод аппроксимации произвольной плотности распределения суммами экспонент // Вестник ВГУ. – 2013. - № 2. - с. 53-57.

ИССЛЕДОВАНИЕ ПЕРИОДИЧЕСКИХ РЕШЕНИЙ В ОКРЕСТНОСТИ ТОЧЕК ЛИБРАЦИИ В КРУГОВОЙ ОГРАНИЧЕННОЙ ЗАДАЧЕ ТРЕХ ТЕЛ

Гафурова Д.Р.

Национальный исследовательский университет «Высшая школа экономики», департамент прикладной математики МИЭМ НИУ ВШЭ

Аннотация

В работе исследуются периодические решения круговой ограниченной задачи трех тел около точек либрации. Для расчета орбит применялся численный

метод, который основан на использовании сечения Пуанкаре периодических орбит. Были получены начальные условия вертикальных и горизонтальных орбит Ляпунова и периодических квазигоризонтальных орбит около точек либрации L1 и L2 в системе Земля-Луна. С помощью анализа множителей Флоке была установлена устойчивость орбит и определены точки бифуркации семейств.

Введение

Решения круговой ограниченной задачи трех тел [1] описывают возможные движения космического аппарата в гравитационном поле двух массивных тел, которые перемещаются по круговым орбитам около их центра масс, например, Солнца и Земли или Земли и Луны. Периодические орбиты около стационарных точек этой задачи, или точек либрации, могут быть полезны при планировании космических миссий. На орбитах около точки либрации L1 в системе Солнце-Земля действовали космические аппараты ISEE-3, SOHO, DSCOVR, Genesis [2], к гало орбите около точки L2 была отправлена миссия Спектр-РГ [3].

Исследуемая задача не имеет общего аналитического решения, и для нахождения её решений разработаны различные численные подходы. В работах [4]–[7] расчет периодических орбит около точек либрации производился методом дифференциальных коррекций. В частности, в работах [5], [6] проведен анализ n -периодических орбит, бифурцирующих из гало орбит, а в [7] – n -периодических квазигоризонтальных орбит ($n = 2, 3$). В работе [8] с использованием метода дифференциальных коррекций был составлен обзор периодических орбит около меньшего из двух массивных тел в системе Юпитер-Европа, который включает в себя орбиты около коллинеарных точек либрации. Аналогичный результат для системы Земля-Луна был получен в [9].

В [10]–[12] представлена методика для расчета орбит около точек либрации, которая основана на использовании геометрического критерия. Такой способ позволяет получать начальные условия периодических и квазипериодических орбит, но только в ограниченной области вокруг точек либрации. Этим подходом были найдены орбиты из основных семейств периодических решений около точек либрации, а также квазигало орбиты, орбиты Лиссажу и квазигоризонтальные орбиты.

В данной работе расчет орбит около точек либрации осуществляется с помощью численного метода, представленного в [13]. Были рассчитаны вертикальные и горизонтальные орбиты Ляпунова около точек либрации L1 и L2 в системе Земля-Луна, и n -периодические квазигоризонтальные орбиты не только для $n = 2, 3$, как в [7], но и для $n = 4, 6$. Результаты исследования выходят за пределы описанной в [12] области применимости.

Круговая ограниченной задачи трех тел

В круговой ограниченной задаче трёх тел описывается движение материальной точки P в гравитационном поле двух массивных тел P_1, P_2 (с массами

$m_1 > m_2$), движущихся по круговым орбитам вокруг их общего центра масс. Для описания движения малого тела вводится подвижная система координат, в которой P_1 и P_2 неподвижны. Началом системы координат является барицентр массивных тел. Ось x совпадает с прямой, соединяющей P_1 и P_2 , и направлена в сторону P_2 . Ось y ортогональна оси x и лежит в плоскости вращения P_1 и P_2 , а ось z дополняет эту систему до правой тройки. В этой системе координат движение малого тела описывается системой дифференциальных уравнений:

$$\begin{cases} \ddot{x} - 2\dot{y} = x - \frac{(1-\mu)(x+\mu)}{r_1^3} - \frac{\mu(x-1+\mu)}{r_2^3} \\ \ddot{y} + 2\dot{x} = y - \frac{(1-\mu)y}{r_1^3} - \frac{\mu y}{r_2^3} \\ \ddot{z} = -\frac{(1-\mu)z}{r_1^3} - \frac{\mu z}{r_2^3}, \end{cases} \quad (1)$$

$$\text{где } \mu = \frac{m_2}{m_1 + m_2}, r_1 = \sqrt{(x - \mu)^2 + y^2 + z^2}, r_2 = \sqrt{(x - 1 + \mu)^2 + y^2 + z^2}.$$

Уравнения (1) можно также представить в виде системы из шести дифференциальных уравнений первого порядка. В данной работе рассматривается система, в которой массивные тела – это Земля и Луна, параметр массы системы $\mu \approx 0.012158$.

Система уравнений (1) имеет пять стационарных решений, которые называются точками либрации. Коллинеарными точками либрации называют три точки, которые принадлежат прямой, соединяющей массивные тела (L1, L2, L3). В окрестности точек либрации существуют семейства периодических орбит. Они могут обладать одним из трех типов симметрий: симметрией относительно плоскости xz , относительно оси x или относительно плоскости xz и оси x .

Множители Флоке

Множители Флоке, или характеристические множители, используются для анализа устойчивости и бифуркаций семейств периодических орбит. Они определяются как собственные значения матрицы монодромии. Матрица монодромии – это фундаментальная матрица системы уравнений, полученной линеаризацией системы (1) около произвольного решения $\bar{X}^*(t)$, полученная для момента времени равного минимальному положительному периоду орбиты. Пусть $\delta\bar{X}(t)$ – отклонение от решения $\bar{X}^*(t)$, тогда для матрицы монодромии верно следующее соотношение:

$$\delta\bar{X}(T + t_0) = \Phi(T, t_0)\delta\bar{X}(t_0), \quad (2)$$

где t_0 – начальный момент времени, T – период орбиты, $\Phi(T, t_0)$ – матрица монодромии.

Вектор состояния в круговой ограниченной задаче трех тел $\bar{X} = (x, y, z, v_x, v_y, v_z)$ состоит из шести компонент: трех координат и трех скоростей, и матрица монодромии имеет размерность 6×6 . Поэтому периодические решения в этой задаче имеют по шесть множителей Флоке. Два собственных значе-

ния матрицы монодромии равны единице, они отображают периодичность орбиты и её принадлежность семейству решений. Другие четыре собственных значения являются парами комплексно-сопряженных чисел и взаимобратных чисел.

Если у периодического решения все множители Флоке по модулю равны единице $|\lambda_i| = 1, i = 1, \dots, 6$, то оно считается устойчивым. Периодическое решение неустойчиво, если у него есть характеристические множители, не принадлежащие единичной окружности, $|\lambda_i| \neq 1$. Так как собственные числа матрицы монодромии - взаимобратные числа, то любому множителю Флоке с модулем больше единицы соответствует множитель Флоке с модулем меньше единицы.

Характеристические множители позволяют обнаружить бифуркации в семействе орбит. Бифуркацией называют качественное изменение решения, вызванное малым изменением начальных условий. В таком случае может возникать семейство периодических орбит, обладающее новыми свойствами. Например, из семейства горизонтальных орбит Ляпунова, которые являются плоскими, бифурцируют трёхмерные гало орбиты.

Определяют несколько типов бифуркаций. При касательной бифуркации комплексно-сопряженные множители Флоке с единичной окружности переходят на вещественную ось, проходя через значение $+1, \lambda_i = \lambda_j = +1$. Бифуркация типа умножение периода на n происходит, когда характеристические множители удовлетворяют соотношению $\lambda^n = 1, n > 1$.

Для нахождения точек бифуркации можно изучать изменение не самих множителей Флоке, а их аргументов при представлении в показательной форме $\lambda = |\lambda|e^{i\varphi}$. Касательная бифуркация происходит, если в семействе решений пара аргументов характеристических множителей сначала уменьшается по модулю, а затем становится равной нулю. Бифуркацию умножение периода на n можно обнаружить у такого решения, у которого есть множители Флоке по модулю равные единице и с показателями $\frac{2\pi}{n}$.

Анализ периодических орбит в окрестности точек либрации

Из точек либрации бифурцируют два типа периодических решений: горизонтальные и вертикальные орбиты Ляпунова. Эти орбиты в окрестности окололунных точек либрации (L1, L2) были исследованы в данной работе, примеры орбит из этих семейств изображены на рис.1 и рис.2.

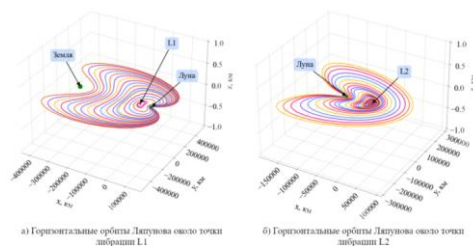


Рис.1. Горизонтальные орбиты Ляпунова

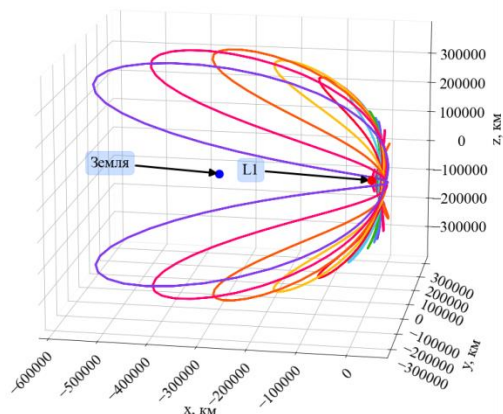


Рис.2. Вертикальные орбиты Ляпунова, бифурцирующие из точки L1

Анализ множителей Флоке также позволил определить, что горизонтальные и вертикальные орбиты Ляпунова вокруг L1, L2 являются неустойчивыми, то есть каждая такая орбита имеет как минимум два характеристических множителя, абсолютные значения которых не равны единице.

Аргументы множителей Флоке горизонтальных орбит Ляпунова представлены на рис. 3 и рис. 4. Из горизонтальных орбит Ляпунова касательной бифуркацией возникают семейства гало орбит и аксиальных орбит. Соответствующие точки бифуркации отмечены оранжевыми и розовыми шестиугольниками. Семейства гало орбит около точек L1 и L2 изучались ранее в [13]. Вертикальные орбиты Ляпунова тоже имеют касательную бифуркацию с аксиальными орбитами.

Горизонтальные орбиты Ляпунова имеют решения, в которых показатели характеристических множителей равны $\pi, \frac{2\pi}{3}$ и $\frac{\pi}{2}$, что свидетельствует о том, что из этих семейств бифурцируют по одному семейству двухпериодических, трёхпериодических и четырёхпериодических орбит.

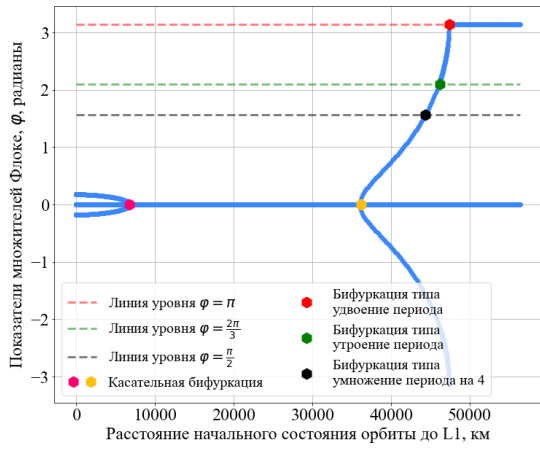


Рис.3. Аргументы множителей Флоке горизонтальных орбит Ляпунова около точки L1

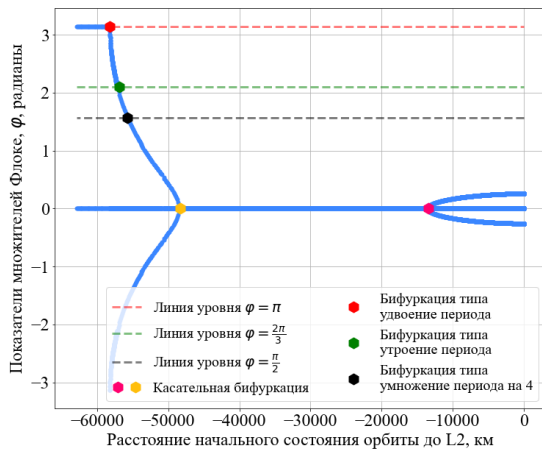


Рис.4. Аргументы множителей Флоке горизонтальных орбит Ляпунова около точки L2

Семейства n-периодических квазигоризонтальных орбит тоже были рассчитаны и изучены в данной работе, их начальные условия отображены на рис.5 и на рис.6.

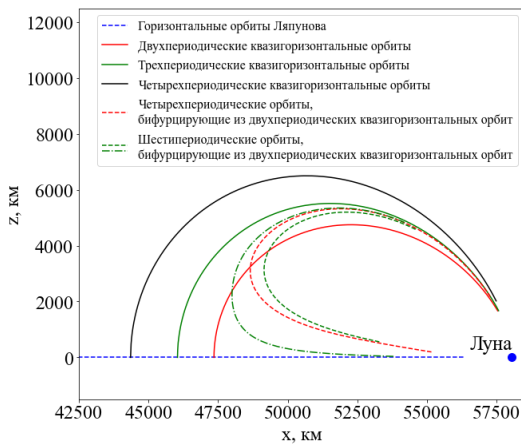


Рис.5. Начальные условия периодических орбит около точки либрации L1

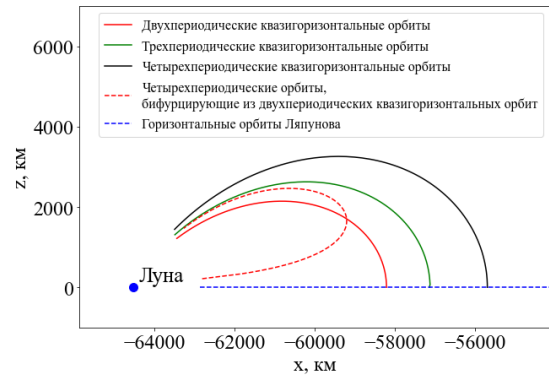


Рис.6. Начальные условия периодических орбит около точки либрации L2

Анализируемые двухпериодические орбиты являются неустойчивыми. Эти семейства имеют по одной точке бифуркации типа удвоение периода, по две точки бифуркации типа утроение периода и по две точки бифуркации типа умножение периода на четыре (рис.7 и рис.8).

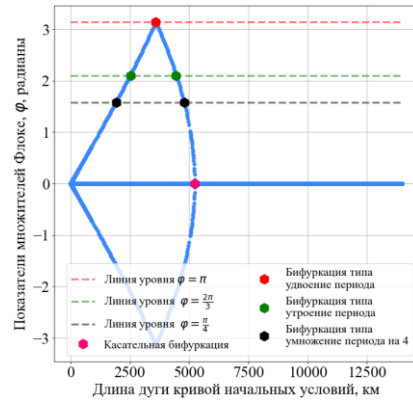


Рис.7. Аргументы множителей Флоке двухпериодических квазигоризонтальных орбит около точки L1

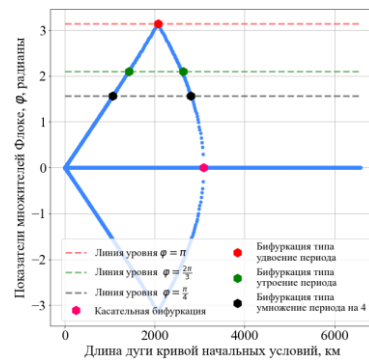


Рис.8. Аргументы множителей Флоке двухпериодических квазигоризонтальных орбит около точки L2

Двухпериодические и четырехпериодические квазигоризонтальные орбиты так же, как и горизонтальные орбиты Ляпунова, из которых они бифурцируют, обладают двумя типами симметрии: относительно плоскости xz и относительно оси x. Однако трехпериодические орбиты, образованные бифуркацией типа

утроение периода из горизонтальных орбит Ляпунова, симметричны только относительно плоскости xz .

Четырехпериодические орбиты, которые бифурцируют из двухпериодических квазигоризонтальных орбит, также стали предметом данного исследования. Их начальные условия отмечены красными пунктирными линиями на рис.5 и рис.6. Примеры двухпериодической орбиты и орбиты из бифурцирующего семейства около точки либрации $L2$ представлены на рис.9. Четырехпериодические орбиты, образованные бифуркацией типа удвоение периода, в отличие от двухпериодических орбит, из которых они бифурцируют, симметричны только относительно плоскости xz .

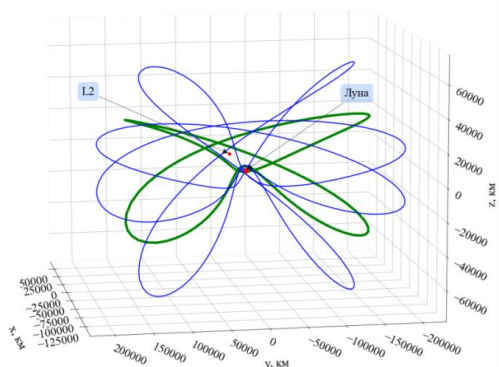


Рис.9. Двухпериодическая квазигоризонтальная орбита (обозначена зеленым цветом) и получаемая из неё бифуркацией удвоение периода четырехпериодическая орбита (обозначена синим цветом) около точки $L2$

Зелеными пунктирной и штрихпунктирной линиями на рис.5 обозначены шестипериодические орбиты, которые бифурцируют из двухпериодических орбит вокруг точки $L1$. Соответствующие точки бифуркации на рис.7 обозначены зелеными шестигульниками. Рис.10 и рис.11 иллюстрируют эти орбиты. Шестипериодические орбиты, отмеченные штрихпунктирной линией, аналогично четырехпериодическим орбитам, при бифуркации из двухпериодических орбит теряют симметрию относительно оси x (рис.10). Орбиты из семейства, обозначенного пунктирной линией, симметричны и относительно оси x , и относительно плоскости xz (рис.11).

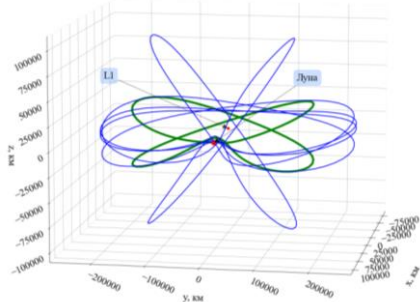


Рис.10. Двухпериодическая квазигоризонтальная орбита (обозначена зеленым цветом) и получаемая из неё шестипериодическая орбита (обозначена синим цветом), в семействе, отмеченном штрихпунктирной линией около точки $L1$

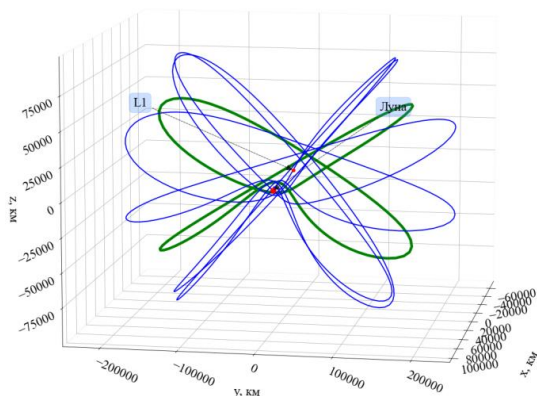


Рис.11. Двухпериодическая квазигоризонтальная орбита (обозначена зеленым цветом) и получаемая из неё шестипериодическая орбита (обозначена синим цветом), в семействе, отмеченном пунктирной линией около точки $L1$

Заключение

В работе были рассчитаны и проанализированы вертикальные орбиты Ляпунова, горизонтальные орбиты Ляпунова, n -периодические семейства орбит, бифурцирующие из горизонтальных орбит Ляпунова бифуркацией умножение периода на n для $n = 2, 3, 4$, а также вторичные бифуркации двухпериодических орбит. В дальнейшем будут построены и изучены семейства n -периодических квазигоризонтальных орбит для $n > 6$ и составлен системный обзор этих решений. Такая работа будет полезна для анализа потенциальных траекторий космического аппарата при планировании космических миссий к точкам либрации.

Список литературы

1. Мюррей К., Дермотт С. Динамика Солнечной системы / Пер. с англ. под ред. И. И. Шевченко. - М.: ФИЗМАТЛИТ, 2010. - 588 с. - ISBN 978-5-9221-1121-8.
2. "NASA, Missions." <https://www.nasa.gov/missions> (дата обращения 15.02.2024)
3. "Российская космическая обсерватория «Спектр-РГ»", URL: <https://www.roscosmos.ru/srg/> (дата обращения 15.02.2024)
4. K. Connor Howell, "Three-dimensional, periodic, halo orbits," *Celest. Mech.*, vol. 32, no. 1, pp. 53–71, Jan. 1984, doi: 10.1007/BF01358403.
5. E. T. Campbell, "Bifurcations from families of periodic solutions in the circular restricted problem with application to trajectory design," Purdue University ProQuest Dissertations Publishing, 1999.
6. E. M. Zimovan-Spreen, K. C. Howell, and D. C. Davis, "Dynamical Structures Nearby NRHOs with Applications to Transfer Design in Cislunar Space," *J. Astronaut. Sci.*, vol. 69, no. 3, pp. 718–744, 2022, doi: 10.1007/s40295-022-00320-4.
7. L. Bury, J. McMahon, and M. Lo, "A study of periodic orbits near Europa," *Celest. Mech. Dyn. Astron.*, vol. 134, no. 3, Jun. 2022, doi: 10.1007/s10569-022-10076-6

8. R. P. Russell, "Global search for planar and three-dimensional periodic orbits near Europa," *J. Astronaut. Sci.*, vol. 54, no. 2, pp. 199–226, Jun. 2006, doi: 10.1007/BF03256483.

9. C. J. Franz and R. P. Russell, "Database of Planar and Three-Dimensional Periodic Orbits and Families Near the Moon," *J. Astronaut. Sci.*, vol. 69, no. 6, pp. 1573–1612, Dec. 2022, doi: 10.1007/s40295-022-00361-9.

10. Y. Ren and J. Shan, "A novel algorithm for generating libration point orbits about the collinear points," *Celest. Mech. Dyn. Astron.*, vol. 120, no. 1, pp. 57–75, 2014, doi: 10.1007/s10569-014-9560-9.

11. H. Zhang and S. Li, "A general method for the generation and extension of collinear libration point orbits," *Celest. Mech. Dyn. Astron.*, vol. 126, no. 4, 2016, doi: 10.1007/s10569-016-9698-8.

12. S. Aksenov, S. Bober, and M. Guskova, "Mapping of initial conditions for libration point orbits," *Adv. Sp. Res.*, vol. 68, no. 6, pp. 2501–2514, 2021, doi: 10.1016/j.asr.2021.04.035.

13. Д. Р. Гафурова, "Построение и исследование периодических решений круговой ограниченной задачи трех тел// Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов имени Е.В. Арменского.," 2023, pp. 18–22.

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ВЫЧИСЛЕНИЯ ОРБИТ ВОКРУГ ТОЧЕК ЛИБРАЦИИ В КРУГОВОЙ ОГРАНИЧЕННОЙ ЗАДАЧЕ ТРЕХ ТЕЛ

Посынкин А.А., Протасов Т.Е.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В работе исследуются орбиты вокруг точки либрации L1 в системе Солнце-Земля (С-3) в круговой ограниченной задаче трех тел. Целью работы являлась разработка и реализация нового способа вычисления орбит и начальных условий, приводящих к ним, с применением машинного обучения.

Введение

Орбиты вокруг точек либрации используются для проектирования миссий, решающих различные задачи. Для изучения солнечного ветра была осуществлена миссия ISEE-3 к точке L1 [1]. Миссия SOHO [2] была направлена на изучение солнечной активности. Целью миссии ACE [3] было изучение частиц солнечного ветра, межзвездной и межпланетной среды. В этих миссиях использовалась возможность непрерывного наблюдения за Солнцем с орбиты вокруг точки либрации.

На данной миссии к точкам либрации остаются актуальными. Так, телескоп «Джеймс Уэбб» [4] был выведен на орбиту вокруг L2 в системе С-3. Такое

расположение позволяет выполнять задачи без помех со стороны Солнца.

На данный момент существует множество методов расчета орбит вокруг точек либрации. К ним относятся методы бисекции, которые применяют разделение фазового пространства на две области [5,6,7], К одному из таких методов относится метод плоскостей, который имеет ограниченную область применимости [7]. Взяв эту область за основу для составления обучающей выборки, планируется создать модель машинного обучения, которая позволит разработать метод, выходящий за границы области работы метода плоскостей.

Постановка задачи

Движение тела в рамках круговой ограниченной задаче трех тел описывается уравнениями (1-3):

$$\begin{cases} \ddot{x} - 2\dot{y} = x - \frac{(x + \mu)(1 - \mu)}{r_1^3} - \frac{\mu(\mu + x - 1)}{r_2^3}, \\ \ddot{y} + 2\dot{x} = y - \frac{(1 - \mu)y}{r_1^2} - \frac{\mu y}{r_2^3}, \\ \ddot{z} = -\frac{(1 - \mu)z}{r_1^2} - \frac{\mu z}{r_2^3}, \end{cases} \quad (1)$$

$$r_1 = \sqrt{(x + \mu)^2 + y^2 + z^2}, \quad (2)$$

$$r_2 = \sqrt{(x - 1 + \mu)^2 + y^2 + z^2}, \quad (3)$$

где μ – константа, вычисляемая для каждой системы.

Первый интеграл системы, называемый константой Якоби, приведен в формуле (4):

$$C = 2U - \dot{x}^2 - \dot{y}^2 - \dot{z}^2 = 2U - v^2, \quad (4)$$

где v – модуль скорости.

Стационарные решения задачи являются точки либрации. Если линеаризовать систему около точки либрации, то решение принимает вид (5):

$$\begin{cases} x = \gamma + A_1 e^{\lambda t} + A_2 e^{-\lambda t} + A_x \cos(\omega_x t + \phi_x) \\ y = cA_1 e^{\lambda t} - cA_2 e^{-\lambda t} + kA_x \sin(\omega_x t + \phi_x) \\ z = A_z \cos(vt + \psi) \end{cases} \quad (5)$$

При $A_1 \neq 0$ решение (траектория) не будет устойчивым а знак A_1 позволит разделить фазовое пространство на две области, а решения – на два класса, соответствующих положительной и отрицательной неустойчивой компоненте. В силу непрерывности пространства на границе между этими областями будут находиться те решения, у которых неустойчивая компонента равна нулю, они соответствуют орбитам вокруг точки либрации. Таким свойством обладает и исходная задача.

Метод плоскостей

Для поиска точки с нулевой неустойчивой компонентой достаточно иметь две точки из разных об-

ластей фазового пространства. Начальное условие, приводящее к орбите, будет лежать между ними. Это свойство используется в исследованиях [5], [6] с применением метода бисекции.

Метод плоскостей [7] позволяет находить начальные условия, приводящие к орбитам, и вычислять сами орбиты, ортогонально пересекающие плоскость $y = 0$. Он применяет классификацию траекторий по знаку неустойчивой компоненты.

Из заданной точки в пространстве со скоростью, ортогональной плоскости $y = 0$, численно интегрируется траектория до достижения координатой x заданного значения x_{min} (плоскости слева) или x_{max} (плоскости справа). Происходит классификация: если достигается левая плоскость, решение относится к отрицательному классу, если правая плоскость – к положительному. Зафиксировав направление скорости и найдя два начальных условия вдоль него с разными знаками неустойчивости, применяется метод бисекции для вычисления начального условия с нулевой неустойчивой компонентой.

Ввиду неустойчивости системы и машинной ошибки, при интегрировании такого начального условия появляется неустойчивая компонента. Для ее устранения осуществляются коррекции по тому же принципу, что был описан выше. Орбита считается успешно вычисленной, если средняя величина коррекций по модулю не превышает значения 10^{-11} .

Ограниченность работы метода объясняется тем, что он применим только для начальных условий, которые приводят к орбитам, не пересекающим заданные плоскости. Примеры орбит, полученных на границе работы метода плоскостей на рис.1:

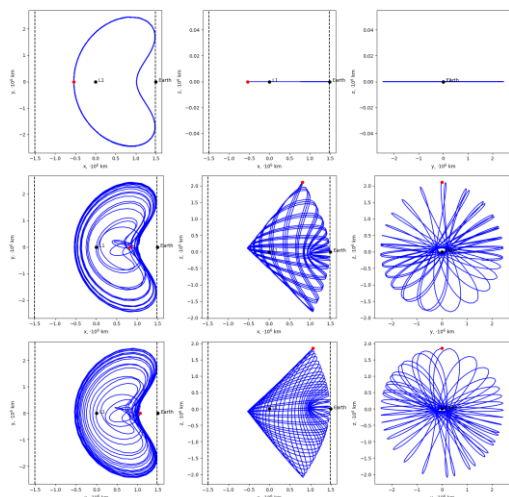


Рис.1. Орбиты в трех проекциях, интегрированные с границы работы метода плоскостей и достигающие правую плоскость (пунктирная линия)

Генерация обучающей выборки

Для генерации обучающей выборки использовался метод обхода линий уровня константы Якоби с применением метода плоскостей. Алгоритм сводится к двум шагам. Первый – вычисление линии нулевых скоростей (ЛНС):

1) Вокруг точки $(x_{L_1}, 0, z_{L_1}, 0, 0, 0)$ задаются начальные условия вида $(x_i, 0, z_i, 0, 0, 0)$, где $x_i = x_{L_1} + r * \cos \alpha_i$, $y_i = y_{L_1} + r * \sin \alpha_i$, r – заданное расстояние, $\alpha_i \in [0, 2\pi]$ – угол инициализации начального условия;

2) Для начальных условий численно интегрируются уравнения (1) и классифицируются методом плоскостей;

3) Осуществляется поиск двух соседних точек с разным знаком неустойчивости;

4) Методом бисекции, варьируя угол инициализации точки, ищется начальное условие с нулевой неустойчивой компонентой;

5) Полученное начальное условие интегрируется на 25 лет с осуществлением коррекции по скорости до 10^{-11} каждые полгода;

6) Пункты (1-5) повторяются для нового начального условия;

7) Пункт (6) выполняется до тех пор, пока не появляется ошибка в интегрировании в пункте (5), то есть, пока не достигается граница работы метода плоскостей.

Второй шаг – обход изолиний. Описание обхода одной изолинии:

1) Для точки ЛНС $(x, 0, z, 0, 0, 0)$ вычисляется константа Якоби C ;

2) Вокруг точки ЛНС строятся начальные условия с $v_y = \pm \sqrt{2U - C}$, $v_y > 0$, если точка «левее» ЛНС, и $v_y < 0$, если «правее»;

3) Далее аналогично пунктам (2-7) из первого шага.

Все траектории, интегрируемые при поиске начальных условий и выполнении коррекций, сохраняются в обучающую выборку.

На рис.2 изображены изолинии константы Якоби, получаемые при генерации выборки:

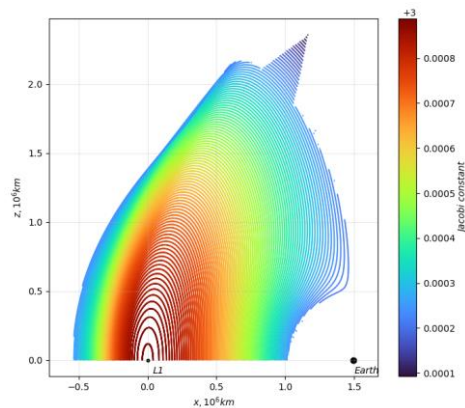


Рис.2. Линии уровня константы Якоби

Разметка траекторий

Чтобы обучить модель, необходимо, чтобы траекториям соответствовали некоторые значения, которые модель будет предсказывать. Эти значения являются разметкой траекторий. Для разметки траекторий сохранялся ее знак и время пересечения плоскости. Для траектории, представляющей из себя множество векторов фазового пространства $\{\vec{x}_i\}$, разметка соответствовала формуле (6):

$$y_i = sign \cdot \exp(2(t_i - \tau)), i = 1..n, \quad (6)$$

где t_i – время, соответствующее x_i , $sign$ – знак траектории, τ – время пересечения плоскости, n – длина траектории.

Такая разметка имеет вид неустойчивой компоненты из системы (5). Пример разметки такого вида на рис.3:

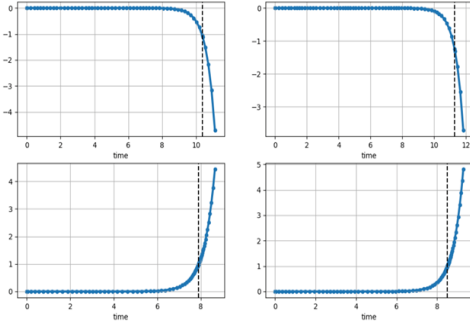


Рис.3. Разметка траекторий по аналогии с неустойчивой компонентой

Модель сможет осуществить классификацию, если предсказанное значение для точки траектории достигает 1 или -1.

К приведенной выше разметке также была применена функция гиперболического тангенса, вследствие чего значения обновленной разметки (7) находились в диапазоне $(-1,0]$ или $[0,1)$:

$$target_i = \tanh y_i, i = 1..n \quad (7)$$

Обновленный способ разметки в сравнении с предыдущим представлен на рис.4:

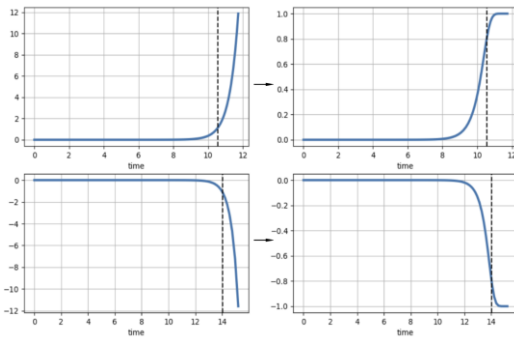


Рис.4. Примеры разметки траекторий новым способом

Здесь классификация осуществляется при достижении значения $\tanh 1 \approx 0.76$.

Модель машинного обучения

Траектории можно интерпретировать как временные ряды, для точек которых нужно решить задачу регрессии, а далее произвести классификацию. Для решения этих задач используется рекуррентная нейронная сеть со слоем LSTM [8]. Его характерным свойством является способность запоминать контекст, давать ответ по текущей точке с учетом

предыдущих. Архитектура слоя LSTM приведена на рис.5:

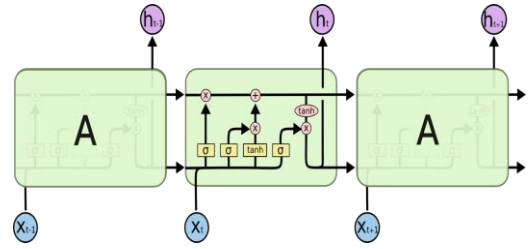


Рис.5. Архитектура LSTM

Модель состоит из нескольких полносвязных слоев до и после LSTM и самих LSTM. Она последовательно принимает точки траектории, предсказывая значение гиперболического тангенса от неустойчивой компоненты до достижения порога отсечки. Для подбора параметров модели сравнивались по количеству успешно найденных начальных условий и соответствующих им орбит вне обучающей выборки. Итоговые параметры модели и их значения:

1. $batch_size = 8$ – число траекторий, подаваемое модели за одну итерацию;
2. $n_layers = 4$ – число полносвязных слоев до и после LSTM, а также в самом LSTM (от 2 до 5);
3. $n_neurons = 128$ – число нейронов в каждом слое;
4. $activation = ReLU$ – функция активации после полносвязных слоев;
5. $learning_rate = 0.001$ – коэффициент скорости обучения;
6. $add_time = 4\pi$ – время, на которое каждая траектория была рассчитана после пересечения плоскости;
7. $delay_time_part = 0.25$ – доля от add_time , на которое был смещен в положительную сторону момент пересечения плоскости;
8. $add_acceleration = False$ – флаг того, что к входным векторам (точкам фазового пространства) добавляется вектор ускорения.

Расширенная карта начальных условий

Итоговая модель была обучена на траекториях, полученных при интегрировании последней изолинии, на которой метод плоскостей работает полностью успешно, ей соответствует $C = 3.000253$. В обучающую выборку вошло 1.1 миллиона траекторий.

Оба шага из раздела «Генерация обучающей выборки» были повторены за пределами работы метода плоскостей с помощью классификации траекторий моделью вместо метода плоскостей. На рис.6 приведена расширенная карта начальных условий в сравнении с методом плоскостей:

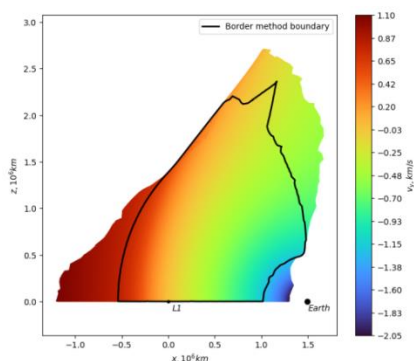


Рис.6. Расширенная карта начальных условий, метод плоскостей применим внутри границ черной кривой

На рис.7 приведен пример орбиты, интегрированной с помощью модели и пересекающей правую плоскость:

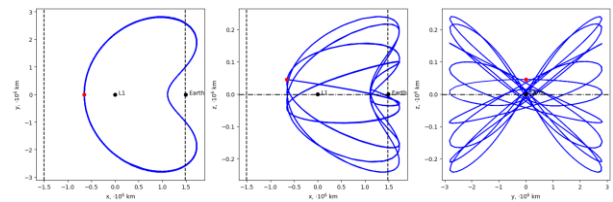


Рис.7. Орбита пересекает плоскость, установленную в рамках метода плоскостей

Заключение

В работе предложен новый метод расчета орбит вокруг точек либрации круговой ограниченной задачи трех тел. Алгоритм расчета орбит основан на классификации траекторий по знаку неустойчивой компоненты посредством машинного обучения. Примененный подход позволил расширить карту начальных условий, полученную методом плоскостей.

Список литературы

1. R. W. Farquhar, “The Flight of ISEE-3/ICE: Origins, Mission History, and a Legacy,” J. Astronaut. Sci., vol. 49, no. 1, pp. 23–73, Mar. 2001, doi: 10.1007/BF03546336.
2. Dunham, D. W. Trajisfer Trajectory Design for the SOHO Libration-Point Mission / D. W. Dunham //IAF Paper 92-0066, September 1992.
3. Stone, E. C. The Advanced Composition Explorer / E.C. Stone // Space Science Reviews, 1998.
4. J. P. Gardner et al., “The James Webb Space Telescope mission,” Publ. Astron. Soc. Pac., vol. 135, no. 1048, p. 068001, 2023
5. M. Hechler and J. Cobos, “Herschel, Planck and Gaia Orbit Design” in Libration Point Orbits and Applications, WORLD SCIENTIFIC, May 2003, pp. 115–135. doi: 10.1142/9789812704849_0006.
6. Y. Ren and J. Shan, “A novel algorithm for generating libration point orbits about the collinear points,”

Celest. Mech. Dyn. Astron., vol. 120, no. 1, pp. 57–75, Sep. 2014, doi: 10.1007/s10569-014-9560-9.

7. S. Aksenov, S. Bober, and M. Guskova, “Mapping of initial conditions for libration point orbits,” Adv. Sp. Res., vol. 68, no. 6, pp. 2501–2514, Sep. 2021, doi: 10.1016/j.asr.2021.04.035.

8. G. Van Houdt, C. Mosquera, and G. Nápoles, “A review on the long short-term memory model,” Artif. Intell. Rev., vol. 53, no. 8, pp. 5929–5955, 2020.

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ НЕЛИНЕЙНЫМ ОБЪЕКТОМ, ЛИНЕАРИЗУЕМЫМ ОБРАТНОЙ СВЯЗЬЮ, В ЗАДАЧЕ ДИФФЕРЕНЦИАЛЬНОЙ ИГРЫ

Гаража И.А.

Национальный исследовательский университет «Высшая школа экономики», департамент прикладной математики МИЭМ НИУ ВШЭ

Аннотация

В работе исследуется задача оптимального управления для нелинейных систем, линеаризуемых обратной связью (ЛОС), рассматриваемая как дифференциальная игра с нулевой суммой [2].

Введение

Для нелинейной системы

$$\dot{x}(t) = f(x(t)) + g_1(x(t))u(t) + g_2(x(t))w(t), \quad (1)$$

для которой существует диффеоморфизм $\Phi(x)$ [1], преобразующий исходную систему с линейной основной частью и нелинейной обратной связью, ставится задача дифференциальной игры двух игроков с нулевой суммой (антагонистическая игра). Система имеет непрерывные функции $f(x)$, $g_1(x)$, $g_2(x)$ и управления $u(t)$, $w(t)$, с ограничениями по мощности, реализующимися с использованием обратной связи по состоянию.

Задача дифференциальной игры как оптимальное управление

Используя предположение об ограничении по мощности: $u^T(t)Ru(t) < E_w$, $w^T Pw < E_w$, $E_u < E_w$, для оценки действий игроков вводится функционал качества (ФК):

$$J = \lim_{t_f \rightarrow \infty} \frac{1}{2} \int_{t_0}^{t_f} \{x^T Qx + u^T Ru - w^T Pw\} dt. \quad (2)$$

Заметим, что значение функционала всегда неотрицательно.

Тогда гамильтониан системы (1) и функционала (2)

$$H = \frac{1}{2} [x^T Qx + u^T Pu - w^T Pw] + \lambda^T [f(x) + g_1(x)u + g_2(x)w]. \quad (3)$$

Тогда, при отсутствии ограничений на управление, из необходимых и достаточных условий минимакса задачи выводится [4] вид управлений:

$$\begin{aligned} u(t) &= -R^{-1} g_1^T(x(t)) \lambda(t), \\ w(t) &= P^{-1} g_2^T(x(t)) \lambda(t). \end{aligned} \quad (4)$$

Вид (4) определяет возможное описание оптимальных управлений. Однако для того, чтобы это управление было оптимальным, необходимо выполнение наложенных ограничений.

Если ввести матрицу

$$\Pi(x(t)) = g_1(x(t)) R^{-1} g_1^T(x(t)) - g_2(x(t)) P^{-1} g_2^T(x(t)),$$

то исходная система преобразуется в

$$\frac{d}{dt} \mathbf{x}(t) = f(\mathbf{x}(t)) - \Pi(\mathbf{x}(t)) \lambda(\mathbf{x}(t)). \quad (5)$$

Если система (5) подлежит ЛОС, т. е. выполняются условия теоремы 5.2.3 [1], то система (5) представима в виде

$$\frac{d}{dt} z(t) = A_0 z(t) + B_0 v(t), \quad (6)$$

где матрицы A_0 и B_0 являются матрицами в форме Бруновского, а $\lambda(t) = \alpha(x(t)) + \beta(x(t))v(x(t))$ - закон управления по каналу обратной связи, то справедливы следующие замечания:

1. При ЛОС ФК существенно меняет свой вид:

$$\begin{aligned} J &= \lim_{t_f \rightarrow \infty} \frac{1}{2} \int_{t_0}^{t_f} \{ (\Phi^{-1}(z))^T(t) Q (\Phi^{-1}(z)) + \\ & [\alpha(\Phi^{-1}(z)) + \beta(\Phi^{-1}(z))v]^T \Pi(\Phi^{-1}(z)) \\ & [\alpha(\Phi^{-1}(z)) + \beta(\Phi^{-1}(z))v] \} dt, \end{aligned} \quad (7)$$

но цели управления остались теми же.

Введём ряд обозначений:

$$\theta(z) = \begin{pmatrix} \frac{\alpha_1(\Phi^{-1}(z))}{n \cdot z_1} & \dots & \frac{\alpha_1(\Phi^{-1}(z))}{n \cdot z_n} \\ \vdots & \ddots & \vdots \\ \frac{\alpha_m(\Phi^{-1}(z))}{n \cdot z_1} & \dots & \frac{\alpha_m(\Phi^{-1}(z))}{n \cdot z_n} \end{pmatrix},$$

$$\zeta(z) = \begin{pmatrix} \frac{\Phi_1^{-1}(z)}{n \cdot z_1} & \dots & \frac{\Phi_1^{-1}(z)}{n \cdot z_n} \\ \vdots & \ddots & \vdots \\ \frac{\Phi_n^{-1}(z)}{n \cdot z_1} & \dots & \frac{\Phi_n^{-1}(z)}{n \cdot z_n} \end{pmatrix}, \quad (8)$$

$$Q(z) = \zeta(z)^T Q \zeta(z) + \theta^T(z) \Pi(\Phi^{-1}(z)) \theta(z),$$

$$N(z) = \theta^T(z) \Pi(\Phi^{-1}(z)) \beta(\Phi^{-1}(z)),$$

$$R(z) = \beta^T(\Phi^{-1}(z)) \Pi(\Phi^{-1}(z)) \beta(\Phi^{-1}(z)).$$

2. Тогда ФК (7) можно преобразовать к ФК с коэффициентами (8), зависящими от состояния

$$\tilde{J} = \lim_{t_f \rightarrow \infty} \frac{1}{2} \int_{t_0}^{t_f} \{ z^T Q(z) z + 2z^T N(z) v + v^T R(z) v \} dt. \quad (9)$$

Переход от уравнения Гамильтона-Якоби-Беллмана к уравнению Риккати с параметрами, зависящими от состояния

Пусть $V^0(t, z(t))$ — функция, непрерывно дифференцируемая по t и $z(t)$ [3], и пара матриц (A_0, B_0) образуют управляемую пару. Тогда для функции $V^0(t, z(t))$, системы (6) и функционала (9) запишем уравнение Гамильтона-Якоби-Беллмана:

$$\inf_{v \in V} \left[\frac{\partial V(t, z)}{\partial t} + \frac{\partial V(t, z)}{\partial z} (A_0 z + B_0 v) + \frac{1}{2} \{ z^T Q(z) z + 2z^T N(z) v + v^T R(z) v \} \right] = 0. \quad (10)$$

Назначим управления в виде

$$v(t) = -R^{-1}(z) \left(B_0^T \left\{ \frac{\partial V(t, z)}{\partial z} \right\}^T + N^T(z) z \right). \quad (11)$$

Таким образом, управление (11) является управлением, организованным по принципу обратной связи, где $\frac{\partial V(t, z)}{\partial z}$ — является решением уравнения

$$\begin{aligned} & \frac{\partial V(t, z)}{\partial z} A_0 z - \frac{1}{2} \frac{\partial V(t, z)}{\partial z} B_0 R^{-1}(z) B_0^T \left\{ \frac{\partial V(t, z)}{\partial z} \right\}^T \\ & + \frac{1}{2} z^T Q(z) z - \frac{1}{2} z^T N(z) R^{-1}(z) B_0^T \left\{ \frac{\partial V(t, z)}{\partial z} \right\}^T - \\ & \frac{1}{2} \frac{\partial V(t, z)}{\partial z} B_0 R^{-1}(z) N^T(z) z - \frac{1}{2} z^T N(z) R^{-1}(z) N^T(z) z = 0. \end{aligned} \quad (12)$$

Для нахождения решения уравнения (12) определим $\frac{\partial V(t, z)}{\partial z}$ как $\frac{\partial V(t, z)}{\partial z} = z^T S^T(z)$. Тогда управление (11) определяется через решение уравнения Риккати с параметрами, зависящими от состояния (SDRE).

Осуществим переход от управления (11) с заменённым $\frac{\partial V(t, z)}{\partial z}$ на $\frac{\partial V(t, z)}{\partial z} = z^T S^T(z)$, найденным $S(z)$ в результате решения SDRE к управлению $\lambda(t) = \alpha(x(t)) + \beta(x(t))v(x(t))$.

В результате ряда преобразований получим

$$\frac{d}{dt} x(t) = f(x(t)) + [\beta^T(x)]^{-1} B_0^T S(\Phi(x)) \Phi(x).$$

Заключение

При введении матрицы Π , найденное оптимальное управление имеет место, если матрица Π является положительно определённой. Данный подход можно расширить, если рассмотреть более полную задачу, при которой ФК будет иметь структуру существенно сложнее текущей.

Список литературы

1. Isidori A. Nonlinear Control Systems. 3rd Edition. - London: Springer, (1995). - 549 p.
2. Афанасьев В.Н. Дифференциальные игры в задачах управления неопределёнными системами. -М.: ЛЕНАРД, (2023). - 400с.
3. Семенов В. В., Пантелеев А. В., Бортакровский А. С. Математическая теория управления в примерах и задачах. — Прикладная математика в примерах и задачах. — МАИ, 1997. — С. 214—216. — ISBN 9785703513941.
4. Афанасьев В.Н., Орлов П.В. “Субоптимальное управление нелинейным объектом, линеаризуемым обратной связью” // Известия РАН. Теория и системы управления. 2011. № 3. С. 14-22.

Секция

«Компьютерные сети, телекоммуникационные системы и интернет вещей»

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ПРОГРАММНО-УПРАВЛЯЕМОГО ИМИТАТОРА ТОКА

Быковский Н.А.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Данная работа посвящена проведению функциональных испытаний малого космического аппарата, а именно систем ориентации и стабилизации на стенде полунатурного моделирования, расположенного в учебно-исследовательской лаборатории на специально разработанном программном обеспечении.

Введение

В условиях быстро прогрессирующего мира интерес людей к исследованию космоса и развитию технологий в этом направлении активно растет. Особенно популярны стали малые космические аппараты, масса которых от нескольких сотен грамм до 500 килограмм. Соответственно, растет спрос на специалистов, оборудование и программное обеспечение, позволяющее проводить предполетные испытания и, тем самым, минимизировать риски возникновения нештатных ситуаций.

Одними из важных систем космического аппарата являются системы ориентации и стабилизации [1, 2]. Благодаря данным системам, космический аппарат стабилизируется относительно центра масс, что необходимо для корректного и безопасного продолжения полета. Для того, чтобы предоставить студентам и сотрудникам университета, занимающимся запуском МКА, возможность бесперебойного проведения экспериментов или функционального тестирования систем аппарата согласно новейшим технологиям было принято решение о разработке аппаратно-программного решения, способного создавать модель полета и траектории МКА на околоземной орбите с помощью алгоритмической модели SGP-4, с возможностью моделирования вектора магнитного поля с помощью модели IGRF-13, и передачей данных вектора на стенд полунатурного моделирования, находящийся в учебно-исследовательской лаборатории.

В данной работе будет рассматриваться способ тестирования систем ориентации и стабилизации малых космических аппаратов и возможности разработанного аппаратно-программного решения.

Описание работы стенда

В центре установки расположен воздушный подшипник (см. рис. 1), состоящий из статора (чаши) со специальными форсунками и ротатора, который служит опорой для установки объекта испытаний.

В статор из воздушного компрессора поступает воздух, благодаря этому происходит формирование тонкого воздушного слоя, который обеспечивает свободное и плавное вращение в трех основных вращательных направлениях с минимальным сопротивлением.

По периметру установки расположена трехкоординатная магнитная система в виде взаимно ортогональных магнитных катушек, которая, с помощью управления током в катушках, позволяет имитировать и изменять магнитное поле в центре магнитной системы. Эта магнитная установка управляется с помощью источника питания – программно-управляемого источника тока (ПУИТ).

В углу установки, на специальном подвесе установлен имитатор солнечного излучения (ИСИ).

ИСИ используется для имитации солнечного излучения.

Имеющийся функционал позволяет проводить испытания в условиях, максимально приближенных к реальным космическим условиям.

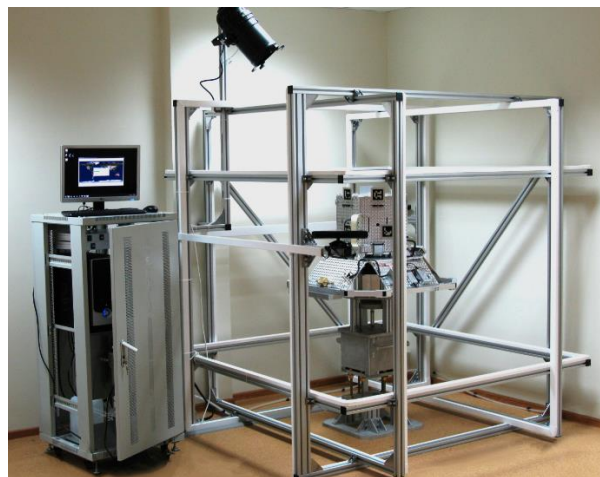


Рис.1. Стенд полунатурного моделирования

Используемые модели

Модель SGP4 [3] (Simplified General Perturbations 4) - это алгоритмическая модель, используемая для прогнозирования и расчета орбитальных параметров искусственных спутников Земли. Она была разработана в рамках программы NORAD (North American Aerospace Defense Command) и является одной из наиболее широко используемых моделей для расчета орбитальных элементов спутников. SGP4 основывается на теории гравитационных воздействий и учитывает действие гравитационных сил от Луны, Солнца, планет и других крупных небесных тел на орбиту спутника. Эта модель также учитывает вращение Земли и атмосферное сопротивление.

Модель IGRF-13, также известная как Международное геомагнитное эталонное поле, представляет собой модель, которая используется для симуляции

геомагнитного поля Земли. Эта модель была разработана и обновляется регулярно с целью обеспечить точное представление изменений и характеристик геомагнитного поля.

Она была разработана Международной службой геомагнитного поля (International Geomagnetic Reference Field Service) и представляет собой последнюю версию этой модели. Модель основана на данных, собранных множеством геомагнитных наблюдательных станций по всему миру. Эти данные включают наблюдения магнитного поля Земли на разных широтах, долготах и глубинах. Используя эти данные, модель IGRF-13 стремится представить геомагнитное поле Земли в любой точке на поверхности Земли и на разных глубинах.

IGRF-13 представляет геомагнитное поле в виде разложения в сферические гармоники. Разложение потенциала поля в ряд, было предложено в 1838 К.Ф. Гауссом. Это разложение имеет вид:

$$V = -R \sum_{i=1}^k \left(\frac{R}{r}\right)^{i+1} \sum_{n=0}^m (g_n^m(t) \cos m\lambda_0 + h_n^m(t) \sin m\lambda_0) P_n^m(\cos \vartheta_0), \quad (1)$$

$$B = \mu_0 \nabla V, \quad (2)$$

Программное обеспечение для ПУИТ

Как было описано ранее, для создания имитации магнитного поля необходимо подавать ток на катушки, которые управляется с помощью программно-управляемого источника тока (ПУИТ). В начале Raspberry Pi [5] по протоколу SPI (Serial Peripheral Interface), который представляет синхронный последовательный интерфейс, позволяя устройствам обмениваться данными в полудуплексном режиме отправляет настройки, в том числе диапазон выходного напряжения на цифроаналоговый преобразователь (ЦАП) от 0 до 2^{16} (так как ЦАП 16-ти битный). Потом это напряжение идёт на усилитель типа АВ и после усилителя ток идёт на катушки. То есть, чтобы создавать магнитное поле, необходимо передавать значение на Raspberry Pi. На этом и строятся основы разработанного программного обеспечения (ПО).

ПО разработано на языке Python [6] с использованием различных фреймворков и библиотек. Оно является кроссплатформенным, что позволяет запускать его на различных ОС (Windows, Linux, macOS). Кроссплатформенность достигается благодаря тому, что Python является интерпретируемым языком, который может быть запущен на разных платформах без необходимости перекомпиляции исходного кода. В основе разработки интерфейса лежит фреймворк PyQt [7].

Разработанное программное обеспечение способно моделировать и анимировать пролет МКА, моделировать вектор магнитного поля и передавать полученные результаты на стенд полунатурного моделирования.

Для моделирования необходимо задать TLE данные [8] и выбрать время моделирования. TLE данные также можно загружать с сервера таких как, Celestrak, NORAD и другие аналогичные.

Во время моделирования моделируются геоцентрические координаты и высота положения над поверхностью земли, согласно SGP-4, в заданном диапазоне времени, по данным координатам, для каждого момента времени моделируется вектор магнитного поля, согласно IGRF-13 и траектория полета. После того, как моделирование завершится, на главном экране программы появится функционал для управления анимацией и отображением орбиты спутника. Во время анимации в верхней части экрана будут отображаться широта долгота и высота (в километрах) текущего положения спутника, а также, вектор магнитного поля в нанотеслах (см. рис. 2).

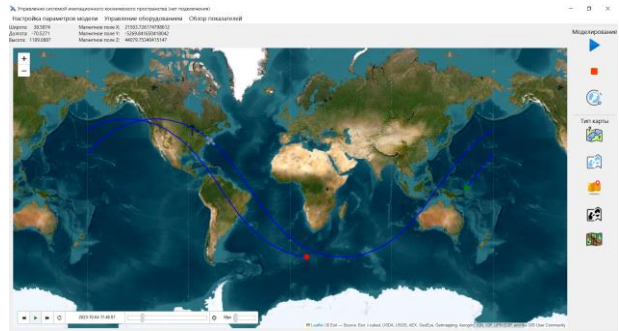


Рис.2. Главное окно программы

Во время анимации пролета малого космического аппарата вектор магнитного поля преобразуется и по локальной сети передается средствами web-сокетов на Raspberry Pi в режиме реального времени. Функционал и последовательность работы программно-аппаратного комплекса представлена на схеме (см. рис. 3).

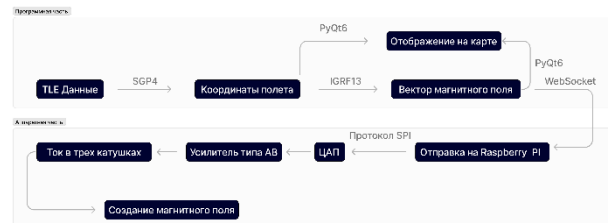


Рис.3. Схема работы программно-аппаратного комплекса

Таким образом испытываются средства ориентации и стабилизации малого космического аппарата, расположенного непосредственно на стенде полунатурного моделирования.

Заключение

Разработанное программное обеспечение обладает возможностями для моделирования пролета МКА, моделирования вектора магнитного поля, а также для изучения и функциональности систем ориентации и стабилизации спутника. Это дает широкие возможности проведения испытаний средств ориентации и стабилизации малых космических аппаратов. Продукт был протестирован в учебно-исследовательской

лаборатории НИУ ВШЭ и может быть использован при планировании запусков следующих спутников.

Список литературы

1. Попов В. И. Системы ориентации и стабилизации космических аппаратов. – 2-е изд., перераб. и доп. – Москва: Машиностроение, 1986. – 184 с.
2. Каргу Л.И. Системы угловой стабилизации космических аппаратов – М.: Машиностроение, 1980. – 172 с.
3. Wei D and Zhao C Y 2010 An Accuracy Analysis of the SGP4/SDP4 Model Chinese Astronomy and Astrophysics 34(1) pp 69-76
4. IAGA V-MOD Geomagnetic field modeling: International Geomagnetic Reference Field. URL: <https://www.ngdc.noaa.gov/IAGA/vmod/igrf.html> (дата обращения: 10.10.2023). – Текст: электронный.
5. Гарет Халфакри. Raspberry Pi. Официальное руководство для начинающих.
6. Дэн Бейдер. Чистый Python. Тонкости программирования для профи. — СПб.: Питер, 2018.
7. Дронов В. А., Прохоренок Н. А. Python 3 и PyQt 6. Разработка приложений, 2022
8. NORAD Two-Line Element Set Format URL: <https://www.celestrak.com/NORAD/elements/> (дата обращения: 19.09.2023). – Текст: электронный.

ВЫЧИСЛЕНИЕ ШИРИНЫ БИСЕКЦИИ ЦИРКУЛЯНТНОГО ГРАФА

Рзаев Э.Р.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Данная работа посвящена исследованию использования параметра ширины бисекции в задаче поиска оптимальных циркулянтных топологий. На основе сгенерированных датасетов проведено исследование целевых параметров циркулянтных графов. Для подкласса циркулянтных графов выведена аналитическая формула подсчета ширины бисекции.

Введение

Разработка и поиск новых топологий для различных практических применений до сих пор является актуальной задачей. Одно из таких направлений – проектирование коммуникационных сетей различного вида – представляет комплексную задачу, которая должна выполняться на всех уровнях OSI [1]. Распространенным является подход, при котором проектирование осуществляется, опираясь на базовую топологию сети. Выбор топологии, которая бы имела оптимальные параметры диаметра, средней длины кратчайших путей (mean path length, MPL), степени вершин, ширины бисекции и т. д., во многом определяет задержку передачи в сети, пропускную способ-

ность, устойчивость к ошибкам узлов сети, затраты ресурсов и т. д. [2].

Циркулянтные графы являются перспективной топологией, которая находит широкое практическое применение в суперкомпьютерных вычислениях [3]– [5], при организации эффективной маршрутизации пакетов в локальных сетях [6]; в компьютерных сетях общего назначения [2], [7], в беспроводных сетях [8] и в сетях на кристалле [9]–[12]. При этом процедура синтеза оптимальных циркулянтов имеет высокую вычислительную сложность и решается методами исчерпывающего поиска [12] со значительными ограничениями даже с использованием суперкомпьютера [5], что определяет необходимость ее дальнейшей оптимизации и ускорения.

Обзор предметной области исследования

Диаметр и среднее расстояние кратчайших путей, а также ширина бисекции относятся к основным параметрам топологии, определяющим эффективность функционирования системы передачи данных. В нашей работе под оптимальным понимается циркулянтный граф, имеющий минимумы диаметра и среднего расстояния при заданных порядке и размерности графа. Решается задача поиска образующих, задающих оптимальный граф.

Одной из проблем синтеза циркулянтов является поиск семейств оптимальных графов с количеством узлов большим ста и степенью вершин большей четырех [13]. Поиску аналитических описаний результатов вычислений статических параметров графов уделяется большое внимание. Так в [14] получена серия параметрически описываемых бесконечных семейств кольцевых циркулянтных сетей степени шесть, что дало возможность поиска общих аналитических методов поиска кратчайших путей циркулянтных графов. В работах [15]–[17] найдены семейства двумерных оптимальных циркулянтов, в [18] – аналитические описывания семейства циркулянтов разных степеней, больших шести. В [19] представлена таблица, позволяющая вычислить диаметр циркулянтных графов класса $C(N; 1, s_1)$ с помощью формул, что также является весомым результатом в области поиска аналитических формул. В [20] представлены формулы вычисления диаметра для серий семейств циркулянтных графов степени шесть.

Аналитические описания семейств оптимальных графов, приведенные в различных работах, не покрывают все возможные варианты оптимальных циркулянтных графов для разных критериев оптимальности и параметров графов. Поиск оптимальных графов для инженерных и научных приложений мы предлагаем осуществлять программными средствами.

Вычисление ширины бисекции

Ширина бисекции графа [5], [21] относится к основным параметрам топологии, определяющим эффективность и надежность функционирования системы. В общем случае алгоритм подсчета ширины бисекции Кернинга-Лина [22] имеет сложность $O(n^2 \log n)$. В силу свойств регулярности и симмет-

ричности циркулярных графов задача выведения формулы подсчета бисекции для сигнатур циркулярных топологий может быть упрощена.

Обозначим показатель бисекции графа G через $BW(G)$. Для полных графов, которые являются циркулярными графами с диаметром $d = 1$, ширина бисекции определяется аналитически: $BW(G) = \binom{N}{2}^2$, если N – четное число, и $BW(G) = \left(\lfloor \frac{N}{2} \rfloor\right)^2 + \left\lfloor \frac{N}{2} \right\rfloor$, если N – нечетное число.

Проведенный нами анализ сгенерированных датасетов показал существование среди оптимальных графов циркулянтов вида $C(N; 1, 2, \dots, \lfloor \frac{N}{2} \rfloor - 1)$ с диаметром равным 2. Графы данного вида имеют монотонно возрастающий ряд образующих, начиная с единицы.

Утверждение 1. Если циркулярный граф G – имеет список образующих, состоящий из монотонно возрастающего ряда натуральных чисел от 1 до $\lfloor \frac{N}{2} \rfloor - 1$, то ширина бисекции $BW(G) = k^2 + k$, где k – количество образующих графа.

В качестве примера рассмотрим граф $C(8; 1, 2, 3)$. На рисунке 1 красным цветом выделены ребра, которые необходимо удалить, чтобы разделить граф на две равнозначные (или наиболее близкие по характеристикам графы, если N – нечетное число).

Ширина бисекции для него $BW(C(8; 1, 2, 3)) = 12$. Количество образующих графа (размерность) $k = 3$. Подставим в формулу $BW(G) = k^2 + k = 3^2 + 3 = 12$.

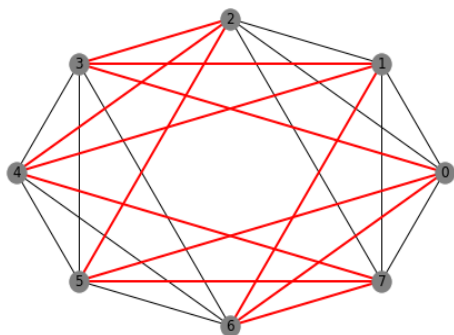


Рис.1. Циркулярный граф $C(8; 1, 2, 3)$

Отметим, что все графы вида $C(N; 1, 2, \dots, \lfloor \frac{N}{2} \rfloor - 1)$ имеют диаметр равный 2, а также являются оптимальными графами, то есть входят в список графов, которые синтезируются с помощью PCG Console [12]. Так как поиск оптимальных графов при больших N может занимать продолжительное время даже на производительном компьютере, то получение аналитических формул для сигнатур графов позволяет сэкономить время поиска подобных графов, а также сузить рассматриваемый спектр подобных сигнатур графов для последующих поисков, что также уменьшает время расчетов.

Рассмотрим уже знакомый граф $C(8; 1, 2, 3)$. Если взять его подграф $C(8; 1, 2)$, который удовлетворяет Утверждению 1, то $BW(C(8; 1, 2)) = k^2 +$

$k = 2^2 + 2 = 6$. Граф $C(8; 1, 2)$ представлен на рисунке 2.

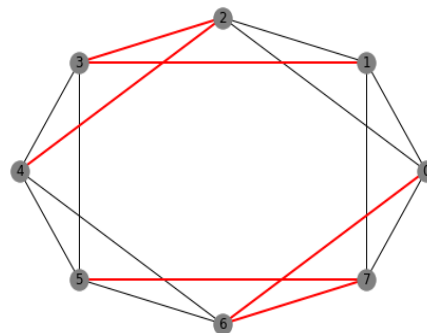


Рис.2. Циркулярный граф $C(8; 1, 2)$

Приведенные расчеты и формулы были получены эмпирическим путем в результате анализа датасета сгенерированных графов и требуют дальнейшего изучения и строгого математического доказательства, которое является предметом для перспективных исследований.

Заключение

В результате проведенного исследования обоснована необходимость расчета ширины бисекции при поиске сигнатур оптимальных циркулярных графов и их анализе. В рамках исследования циркулярных графов выведены аналитические формулы для расчета ширины бисекции для нескольких семейств циркулярных графов.

Список литературы

1. G. Howser, "The OSI Seven Layer Model," in Computer Networks and the Internet, 2020, doi: 10.1007/978-3-030-34496-2_2.
2. H. Liu, X. Li, and S. Wang, "Construction of Dual Optimal Bidirectional Double-Loop Networks for Optimal Routing," Mathematics, 2022, doi: 10.3390/math10214016.
3. C. H. Liang, C. H. Cheng, H. L. Wu, C. C. Li, P. L. Huang, and C. C. Hwang, "Performance evaluation of multi-exaflops machines using Equality network topology," J. Supercomput., 2023, doi: 10.1007/s11227-022-05005-1.
4. F. C. M. Lau and G. Chen, "Optimal layouts of midimew networks," IEEE Trans. Parallel Distrib. Syst., 1996, doi: 10.1109/71.536939.
5. X. Huang, A. F. Ramos, and Y. Deng, "Optimal circulant graphs as low-latency network topologies," J. Supercomput., 2022, doi: 10.1007/s11227-022-04396-5.
6. H. P. Dharmasena and X. Yan, "An optimal fault-tolerant routing algorithm for weighted bidirectional double-loop networks," IEEE Trans. Parallel Distrib. Syst., 2005, doi: 10.1109/TPDS.2005.103.
7. R. Beivide, C. Martínez, C. Izu, J. Gutierrez, J. Á. Gregorio, and J. Miguel-Alonso, "Chordal topologies for interconnection networks," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2003, doi: 10.1007/978-3-540-39707-6_33.

8. A. Mukherjee, P. K. Deb, and S. Misra, "Timed Loops for Distributed Storage in Wireless Networks," *IEEE Trans. Parallel Distrib. Syst.*, 2022, doi: 10.1109/TPDS.2021.3100780.

9. E. A. Monakhova, O. G. Monakhov, and A. Y. Romanov, "Routing Algorithms in Optimal Degree Four Circulant Networks Based on Relative Addressing: Comparative Analysis for Networks-on-Chip," *IEEE Trans. Netw. Sci. Eng.*, 2022, doi: 10.1109/TNSE.2022.3211985.

10. N. H. Attrah, G. H.abdul-Majeed, and M. Z. Abdullah, "Implementation of chordal ring network topology to enhance the performance of wireless broadband network," *EUREKA, Phys. Eng.*, 2021, doi: 10.21303/2461-4262.2021.001734.

11. A. Y. Romanov, N. M. Myachin, E. V. Lezhnev, A. D. Ivannikov, and A. El-Mesady, "Ring-Split: Deadlock-Free Routing Algorithm for Circulant Networks-on-Chip," *Micromachines*, 2023, doi: 10.3390/mi14010141.

12. A. Romanov, "The Dataset for Optimal Circulant Topologies," *Big Data Cogn. Comput.*, vol. 7, no. 2, p. 80, Apr. 2023, doi: 10.3390/bdcc7020080.

13. O. Monahov and E. Monakhova, "Regular Network Class Features Enhancement Using an Evolutionary Synthesis Algorithm," *Sci. Educ. Bauman MSTU*, 2014, doi: 10.7463/1014.0728878.

14. E. A. Monakhova, "A set of families of analytically described triple loop networks defined by a parameter," *Prikl. Diskretn. Mat.*, vol. 0, no. 49, pp. 108–119, 2020, doi: 10.17223/20710410/49/8.

15. J. -C Bermond and D. Tzvieli, "Minimal diameter double-loop networks: Dense optimal families," *Networks*, 1991, doi: 10.1002/net.3230210102.

16. B. X. Chen, J. X. Meng, and W. J. Xiao, "Some new optimal and suboptimal infinite families of undirected double-loop networks," *Discret. Math. Theor. Comput. Sci.*, 2006, doi: 10.46298/dmtcs.377.

17. P. K. Jha, "Dense bipartite circulants and their routing via rectangular twisted torus," *Discret. Appl. Math.*, 2014, doi: 10.1016/j.dam.2013.09.021.

18. R. R. Lewis, "Analysis And Construction Of Extremal Circulant And Other Abelian Cayley Graphs," p. 389, 2021, doi: 10.21954/ou.ro.00013612.

19. L. Loudiki, M. Kchikech, and E. H. Essaky, "A new approach for computing the distance and the diameter in circulant graphs," pp. 1–14, Oct. 2022, doi: 10.48550/arXiv.2210.11116

20. E. A. Monakhova and O. G. Monakhov, "Constructing a Series of Families of Degree Six Circulant Networks," *J. Appl. Ind. Math.*, vol. 16, no. 4, pp. 695–705, Nov. 2022, doi: 10.1134/S199047892204010X.

21. B. Mans and I. Shparlinski, "Random Walks, Bisections and Gossiping in Circulant Graphs," *Algorithmica*, vol. 70, no. 2, pp. 301–325, Oct. 2014, doi: 10.1007/s00453-013-9810-3.

22. T. Bui, C. Heigham, C. Jones, and T. Leighton, "Improving the performance of the Kernighan-Lin and simulated annealing graph bisection algorithms," in *Proceedings – Design Automation Conference*, 1989, doi: 10.1145/74382.74527.

СИСТЕМА ПОЗИЦИОНИРОВАНИЯ ВНУТРИ ПОМЕЩЕНИЙ ДЛЯ ПЛАТФОРМЫ ИНТЕРНЕТА ВЕЩЕЙ

Лясковский А.Д.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Данная работа посвящена анализу алгоритмов позиционирования в пространстве внутри помещений, применимых для облачной платформы интернета вещей. В рамках работы были рассмотрены основные способы позиционирования и было выбрано решение, позволяющее добиться наибольшей точности.

Введение

Одной из актуальных задач, в выполнении которых значительно помогло применения IoT устройств, является позиционирование персонала в пространстве [1]. Определение местоположения людей может быть использовано как на больших производственных предприятиях для оптимизации процессов посредством определения местонахождения персонала, охраны труда, так и в коммерческом секторе для предоставления индивидуальных услуг клиентам в зависимости от их местоположения. Вне помещений проблема позиционирования в большинстве случаев решается за счет глобальных навигационных спутниковых систем (GLONASS, GPS, Galileo), однако внутри помещений мощность спутникового сигнала ослабевает, точности уменьшается [2]. Следовательно, для навигации и позиционирования внутри помещений требуется использование иных методов. Позиционирование в таком случае основывается на использовании элементов внутренней инфраструктуры. Существует ряд методов для организации систем внутреннего позиционирования, основанных на применении различных алгоритмов позиционирования, протоколов и аппаратного обеспечения, но даже сейчас сохраняется потребность в недорогих, но при этом точных системах позиционирования. Цель данной работы – исследовать существующие методы позиционирования, выбрать наиболее пригодный, оценив точность, масштабируемость и издержки выбранного метода.

Технологии локального позиционирования

Одной из самых точных технологий, которые используются для определения местоположения в режиме реального времени является UWB [3]. За счет своей низкой частоты пропускания, она практически не оказывает помех для других типов коммуникации, проведенных в тех же помещениях (например, Wi-Fi), а также характеризуется высоким уровнем помехозащищенности. Однако главными недостатками являются малый радиус действия и относительно высокая цена оборудования по сравнению с систе-

мами позиционирования, построенными на основе других технологий.

Одним из самых распространенных технологий не только в рамках IoT систем, но и в целом, является Wi-Fi [4]. За счет своей распространенности и низкой стоимости оборудования Wi-Fi часто рассматривается как один из потенциальных вариантов для создания системы позиционирования, так как чаще всего это позволяет использовать уже существующую инфраструктуру без привлечения дополнительных затрат для построения отдельной системы позиционирования. В то же время Wi-Fi изначально не предназначался для целей создания систем позиционирования, поэтому загруженность эфира другими устройствами, низкая точность (10-15 м.) и высокое энергопотребление устройствами делают данную технологию не лучшим выбором для создания систем позиционирования.

BLE – это спецификация Bluetooth, предназначенная для маломощных устройств. Технология уже устоялась на рынке беспроводных персональных сетей за счет компактности и дешевизны модулей и за счет использования именно спецификации для устройств с низким энергопотреблением отлично подходит для создания систем позиционирования [5].

Методы позиционирования

Одним из самых распространённых принципов для создания систем позиционирования вне зависимости от выбираемой технологии является создание сети из «якорных» устройств, называемых маяками, и носимых устройств, которые закрепляются за конкретными пользователями. В качестве носимого устройства могут выступать как уже имеющиеся у пользователей устройства (например, смартфоны), так и отдельные. Координаты располагаемых маяков изначально известны и не меняются по мере работы системы, поэтому местоположение пользователя может быть вычислено на основе характеристик сигналов, которое получает носимое устройство от них. В свою очередь, методы позиционирования делятся на группы в зависимости от характеристики сигнала, используемой для определения расстояния [6].

Методы ToF (Time of Flight) и TDoA (Time Difference of Arrival) используют в качестве ключевого параметра для вычисления расстояния между носимым устройством и маяком задержку распространения радиосигнала между ними. Время, затраченное на распространение сигнала, с учетом известности скорости распространения, может быть использовано для вычисления расстояния, прошедшего сигналом до устройства. Несмотря на относительно высокую точность определения местоположения при правильной исходной настройке устройств, данные методы сильно усложняют затраты на обслуживание, так как требуют синхронизации внутреннего времени всех участвующих в обмене сигналами устройств. Даже незначительная погрешность в разнице внутренних таймеров устройств будет иметь фатальный эффект на точность определения местоположения. Чаще всего носимые устройства выводятся из экс-

плуатации и меняются на новые по мере работы системы, каждое новое носимое устройство требует тестирования на синхронизацию времени с маяками, что делает данные методы крайне неудобными и сложными для применения на практике.

Метод AoA (Angle of Arrival) основан на использовании угла поступления сигнала для определения местоположения посредством применения базовых принципов геометрии. Данный метод требует снабжения маяков несколькими антеннами или антенной решеткой для определения вышеупомянутого угла поступления сигнала, что увеличивает затраты на приобретение маяков. Более того, метод AoA имеет высокую точность при правильном расположении маяков в помещении, но точность разительно ухудшается при выходе за пределы рекомендуемого радиуса действия. Это делает AoA слабо применимым для сооружений сложной структуры и больших помещений.

Метод, основанный на использовании RSSI, использует уровень мощности принимаемого сигнала для определения расстояния между маяком и устройством. Данный метод может быть прост на этапе монтажа, так как требует правильного расположения самих маяков – станцию нельзя загромождать перекрывающими распространение сигнала предметами, сами станции должны располагаться на расстоянии не менее 3-х метров друг от друга. Однако он прост в эксплуатации, так как не требует синхронизаций между станциями, потребляет мало энергии и имеет приемлемую точность (до 5 м.), которая также может быть повышена посредством использования алгоритмов позиционирования. За счет вышеописанных достоинств и низкой цены реализации метод, использующий RSSI, является самым распространенным из всех вышеописанных методов.

Алгоритмы позиционирования

Алгоритм ближайшего маяка (Proximity) [7] является самым простым с точки зрения вычислений – координаты пользователя определяются по координатам точки доступа, которая находится к нему ближе всего. Ввиду своей простоты погрешность остается крайне высокой и сильно зависит от плотности расположения точек доступа в помещении.

Более продвинутым алгоритмом является принцип центроида или взвешенного центроида, где положение пользователя определяется как геометрический центр, образованный проекциями распространения сигнала от маяков на одну из плоскостей. В случае взвешенного центроида учитываются не только местоположения самих станций, но и мощности сигналов. В таком случае местоположение определяется по формуле (1):

$$\begin{cases} X_0 = \sum_{i=1}^N w_i X_i \\ Y_0 = \sum_{i=1}^N w_i Y_i \end{cases} \quad w_i = \left(P_i^2 \sum_{j=1}^N \frac{1}{P_j^2} \right)^{-1} \quad (1)$$

где X_0, Y_0 – декартовы координаты расположения устройства пользователя, X_i, Y_i – декартовы координаты i -того маяка, w_i – характеристика мощности сигнала i -того маяка, N – количество маяков, P_i – значение RSSI, полученное от i -того маяка.

Алгоритм трилатерации основан на нахождении расстояния между устройством и как минимум тремя маяками. Для вычисления координат устройства необходимо решить систему уравнений (2):

$$\begin{cases} (X_0, Y_0) = \operatorname{argmin}[\gamma(x, y)] \\ \gamma(x, y) = \sum_{i=2}^N \left| \frac{P_i}{P_1} - \frac{\lg[d_i(x, y)]}{\lg[d_1(x, y)]} \right|^2 \end{cases} \quad (2)$$

Результаты тестирования приведены в таблице 1 и показывают, что среди всех вышеописанных алгоритмов дифференциальная трилатерация имеет наименьшую погрешность и в отличие от других алгоритмов, имеющих сопоставимую точность, не имеет большой зависимости от внешних факторов.

Таблица 1. Результаты тестирования

| Алгоритм | Кол-во маяков | Погрешность (м) | Условия тестирования |
|-------------------------------|---------------|-----------------|---|
| Proximity | 3 | 30,2 | |
| | 4 | 25 | |
| Центроид | 3 | 25,3 | |
| | 4 | 17 | |
| Взвешенный центроид | 3 | 24,6 | |
| | 4 | 12 | |
| Fingerprint | 4 | 12,1 | (10;25), (25;40), (40;25), (25;10), (25;25) |
| | 4 | 3 | (5;5), (5;45), (45;45), (45;5), (25;25) |
| Трилатерация | 4 | 4 | n=3,4 |
| | | 0,2 | n=3,5 |
| | | 10,6 | n=4 |
| Дифференциальная трилатерация | 3 | 8,7 | |
| | 4 | 3,1 | |

Примечание. В таблице 1 в ряде тестов помимо количества маяков менялись другие показатели, влияющие на точность – они представлены в столбце «Условия тестирования». Для метода Fingerprint менялись декартовы координаты маяков, а для метода трилатерации – коэффициент распространения сигнала в среде.

Алгоритм сопоставления с образцом (Fingerprinting) [7] в отличие от других алгоритмов требует ин-

формации о предыдущем состоянии среды, в которой происходит вычисление местоположения, а значит требует дополнительной стадии для изначальной подготовки. Алгоритм работает следующим образом – на этапе подготовки происходит сбор характеристик, получаемых от маяков, с учетом того, что нам известно местоположение конечного устройства. С учетом подготовки на первом этапе, где мы получаем слепки данных, в дальнейшем, при работе на практике, происходит корректировка результата посредством некоторого алгоритма. В качестве алгоритма в данном случае могут выступать функция аппроксимации, нейронные сети [8], фильтр Калмана [9], k-means [10]. Сопоставление с образцом может быть использовано параллельно другим алгоритмам определения местоположения, например, совместно с трилатерацией, для дополнительного улучшения точности местоположения. Исследования показывают [11], что правильно подобранная и натренированная нейронная сеть в совокупности с функцией аппроксимации могут снизить среднюю погрешность в определении местоположения с 3 м. до <1 м., при этом оставляя издержки на аппаратную часть подобной системы крайне низкими по сравнению с конкурентными решениями.

Заключение

В результате проведенного анализа различных технологий и алгоритмов для систем позиционирования была выбрана система позиционирования на основе BLE маяков с использованием алгоритма дифференциальной трилатерации за счет своей дешевизны в обслуживании и монтаже, так и приемлемой точности (3-5 м). Для дополнительного повышения точности были рассмотрены алгоритмы сопоставления с образцом, которые, несмотря на необходимость введения дополнительного, подготовительного шага, могут снизить погрешность в определении местоположения до незначительных величин (меньше 1 м.) за счет использования нейронных сетей и функций аппроксимации. Подобная конфигурация системы обладает низким энергопотреблением, низкой стоимостью и при этом высокой точностью, что делает её одним из самых удобных для использования вариантов при создании систем определения местоположения внутри помещений.

Список литературы

1. T. Mustafa and A. Varol, "Review of the Internet of Things for Healthcare Monitoring," 2020 8th International Symposium on Digital Forensics and Security (ISDFS), Beirut, Lebanon, 2020, pp. 1-6, doi: 10.1109/ISDFS49300.2020.9116305.
2. Bui, V.; Le, N.T.; Vu, T.L.; Nguyen, V.H.; Jang, Y.M. GPS-Based Indoor/Outdoor Detection Scheme Using Machine Learning Techniques. *Appl. Sci.* 2020, 10, 500, doi: 10.3390/app10020500
3. Li, B., Zhao, K. & Sandoval, E.B. A UWB-Based Indoor Positioning System Employing Neural Networks. *J geovis spat anal* 4, 18 (2020). <https://doi.org/10.1007/s41651-020-00059-2>

4. Gentner C. et al. WiFi-RTT indoor positioning //2020 IEEE/ION Position, Location and Navigation Symposium (PLANS). – IEEE, 2020. – С. 1029-1035.
5. Jianyong Z. et al. RSSI based Bluetooth low energy indoor positioning //2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN). – IEEE, 2014. – С. 526-533.
6. Asaad S. M., Maghdid H. S. A comprehensive review of indoor/outdoor localization solutions in IoT era: Research challenges and future perspectives //Computer Networks. – 2022. – Т. 212. – С. 109041.
7. Минахметов Р. М., Рогов А. А., Цымблер М. Л. Обзор алгоритмов локального позиционирования для мобильных устройств //Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. – 2013. – Т. 2. – №. 2. – С. 83-96.
8. Weerasinghe Y. S. P., Maduranga M. W. P., Disanayake M. B. RSSI and feed forward neural network (FFNN) based indoor localization in WSN //2019 national information technology conference (NITC). – IEEE, 2019. – С. 35-40.
9. Lee S. H. et al. Method for improving indoor positioning accuracy using extended Kalman filter //Mobile Information Systems. – 2016. – Т. 2016. doi: 10.1155/2016/2369103
10. Kupper A. Location-Based Services, John Wiley & Sons Ltd, England, 2005, 365 p.
11. Астафьев А. В. и др. Метод позиционирования мобильного устройства с использованием сенсорной сети BLE-маяков, аппроксимации значений уровней сигналов RSSI и искусственных нейронных сетей //Компьютерная оптика. – 2021. – Т. 45. – №. 2. – С. 277-285. doi: 10.18287/2412-6179-CO-826

НАЗЕМНАЯ ОПТИЧЕСКАЯ ПРИЕМНАЯ СИСТЕМА ДЛЯ КВАНТОВОГО КАНАЛА СВЯЗИ СПУТНИК-ЗЕМЛЯ

Бахшалиев Р.М.¹, Мерзлинкин В.Е.²

¹*Национальный исследовательский университет «Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ,
²НИТУ МИСИС*

Аннотация

В работе рассматривается оптическая часть наземного сегмента квантового канала связи для регистрации излучения с малого космического аппарата (далее мка) «Импульс-1», находящегося на низкой околоземной орбите. Данная работа является частью проекта двусторонней оптической квантовой связи Спутник-Земля НИТУ МИСИС, НИУ ВШЭ, Российского квантового центра, компании ЗАО Сконтел и Qspace.

Введение

В настоящее время потребность человечества в передаче большого объема информации растет с

каждым годом, потому как увеличивается и сам объем требуемой для организации современной жизни человека информации. К такой информации может относиться, например, банковская информация, медиа, новостная, правительственная и др. и зачастую эту информацию требуется передавать на большие расстояния, не ограничиваясь пределами города или даже страны.

Задача разработки квантовой связи в открытом пространстве между наземным оптическим терминалом и мка является очень востребованной и актуальной, поскольку такой физический способ передачи информации имеет ряд преимуществ перед существующими в радиодиапазоне. Так, канал оптической связи имеет существенно меньшие потери за счет наиболее узкой направленности диаграммы излучения в силу дифракционной природы электромагнитного излучения. Эта же узко направленность дополнительно является фактором увеличения скрытности и снижения вероятности перехвата информации. Оптический канал связи является наиболее помехозащищенным ввиду плотной загруженности эфира используемыми радиочастотами. Также большая часть радиодиапазона регламентировано закрыта для повсеместного использования. Самым главным преимуществом оптических линий связи является достижение потенциально больших скоростей передачи, что служит основным фактором развития направления для все возрастающей потребности по передаче большого объема информации в единицу времени.

В данной работе рассматривается построение оптической схемы наземной приемной станции для согласованной работы с приемно-передающим каналом, расположенным на мка Импульс-1.

Вводные данные для проектирования и расчет оптической схемы

Суть работы канала связи заключается в следующем: на солнечно-синхронной орбите Земли находится мка Импульс-1 формата «кубсат», разработанной компанией Qspace совместно с НИТУ МИСИС и запущенный в июне 2023 года с космодрома Восточный. Полезная нагрузка аппарата (далее ПН) оснащена приемно-передающим лазерным модулем с гибкой логикой работы [1]. ПН предназначена для работы с наземными оптическими станциями по атмосферному каналу связи. Приемник и передатчик как мка, так и наземной станции оснащены оптическими антеннами – телескопами в упрощенном смысле этого слова (в данной работе рассматривается оптическая часть приемного канала наземной станции) [2]. Эти антенны служат для пространственного преобразования оптического излучения. В качестве передатчика и приемника используются как полупроводниковые, так и сверхпроводящие квантовые устройства. В процессе работы для обеспечения «видимости» передатчика приемником необходимо выполнять условие соосности двух систем с заданной точностью.

Задача проектирования двустороннего канала связи является комплексной, базирующейся на ре-

шении различных научных и технических вопросов, но как основа, для проектирования именно оптической части, выступает решение так называемого энергетического уравнения системы, которое отражает в себе принципиальную возможность работы связки передатчик – приемник в условиях различных физических ограничений, то есть по сути, решение должно удовлетворять выбранному соотношению сигнал/шум. В общем основная задача оптической схемы состоит в том, чтобы реализовать минимально возможное соотношение сигнал/шум при условии максимальной дальности связи. Дополнительно успешное решение такого уравнения может снизить энергозатраты как передатчика, так и приемника космического сегмента, где требования к таким параметрам – высоки. Возможно по-разному комбинировать параметры в уравнении для достижения цели, но в нашем случае строго зафиксированными параметрами являются динамические диапазоны работы квантовых устройств в большей степени на мка, учитывая требования на энергообеспечение. Рассмотрение построения оптической схемы мка выходит за рамки этой работы, ограничимся лишь табличными характеристиками канала.

Таблица 1. Основные характеристики ПН Импульс-1

| Характеристика | Значение |
|--|----------|
| Длина волны передающего канала, нм | 852 |
| Расходимость передающего канала, мкрад | 160 |
| Оптическая мощность приемного канала, Дб | 20 |
| Длина волны приемного канала, нм | 808 |
| Поле зрения приемного канала, град | ±1,5 |
| Минимальная чувствительность, Дб | -45 |

В общем виде упрощенное энергетическое уравнение, применительное к описанной системе имеет вид:

$$K = \frac{P_{пр}}{P_{пер}} \approx \frac{k_{пер} \cdot k_{атм} \cdot k_{пр} \cdot D_2^2}{4 \cdot \beta^2 \cdot L^2} \quad (1)$$

Коэффициент ослабления излучения равен отношению мощности излучения на входе чувствительной площадки приемного канала ($P_{пр.}$), к мощности излучателя передающего канала ($P_{пер.}$)

$k_{пер.}$ – коэффициент использования излучения в передающем канале.

$k_{атм.}$ – коэффициент пропускания атмосферного канала.

$k_{пр.}$ – коэффициент использования излучения в приемнике.

$k_{атм.}$ – коэффициент пропускания слоя атмосферы

D_2 – диаметр входного зрачка приемника.

β – половина угла расходимости излучения на выходе оптической схемы мка

L – расстояние между приемником и передатчиком

Далее, зная оптическую мощность передатчика можно построить совокупность решений уравнения (1) в зависимости от таких факторов как ослабление в среде распространения и расстояния между приемником и передатчиком. Как известно относительно наземной станции мка движется по закону небесного движения согласно параметров его орбиты и были построены аналитические зависимости изменения расстояния с течением времени сеанса связи для различных наклонений траекторий движения мка [3]. Далее был проделан широкий пласт работы по оценке коэффициента пропускания атмосферного канала, который как известно зависит от факторов рассеяния, молекулярного поглощения, турбулентности приземного слоя – все это уменьшает коэффициент пропускания. Модель аэрозольного и турбулентного ослабления выглядят следующим образом:

$$\tau_{атм}^{аэроз} = e^{(-\varepsilon \cdot L)} \quad (2)$$

$$\varepsilon(\lambda) = \frac{3,91}{S_M} \cdot \left(\frac{0,55}{\lambda}\right)^{0,585 \cdot \sqrt{S_M}} \quad (3)$$

где S_M – метеорологическая дальность видимости в зависимости от характеристики видимости (табличная величина), λ – длина волны излучения, L – протяженность трассы аэрозольного ослабления, ε – спектральное ослабление.

$$\tau_{атм}^{тур} = e^{-K_T \cdot L} \quad (4)$$

$$K_T = 0,4 \cdot k_\lambda^2 \cdot C_n^2 \cdot d^{\frac{5}{3}} \quad (5)$$

$$k_\lambda = \frac{2 \cdot \pi}{\lambda} \quad (6)$$

где K_T – коэффициент ослабления, L – трасса распространения турбулентности, k_λ – волновое число, C_n^2 – масштаб турбулентности, d – диаметр приемной апертуры.

Были построены математические модели, учитывающие влияние этих факторов. Таким образом появилось конечное количество решений энергетического уравнения, хотя и число таких решений достаточно велико. На основании этих решений, конечно, учитывая весовые функции каждой из составляющих появились первые характеристики оптической схемы наземного канала.

Помимо получения энергетики немаловажным фактором является поле зрения оптического канала, которое позволяет «видеть» излучение при различных наклонах наземной станции относительно прямого направления на мка. Совершенно очевидно, что малое поле зрения предъявляет высокие требования к точности наведения наземной станции на мка, а большое поле зрения приводит к увеличению шума в приемной части. Поэтому задача носит оптимизационный характер и была успешно решена для выбранной конфигурации, благодаря работе подвижного устройства – монтировки, обеспечивающего высокую точность наведения и ведения направления на мка.

В качестве базы для оптической схемы наземной станции была выбрана схема зеркального телескопа

Ричи-Кретъена [4] с диаметром главного зеркала 600 мм, с относительным отверстием 1:8. Для обеспечения работы канала связи, данная схема была доработана и дополнена оптическими элементами, расчет которых был произведен в специализированной САПР «Zemax» [5]. Вид схемы показан на рис. 1.

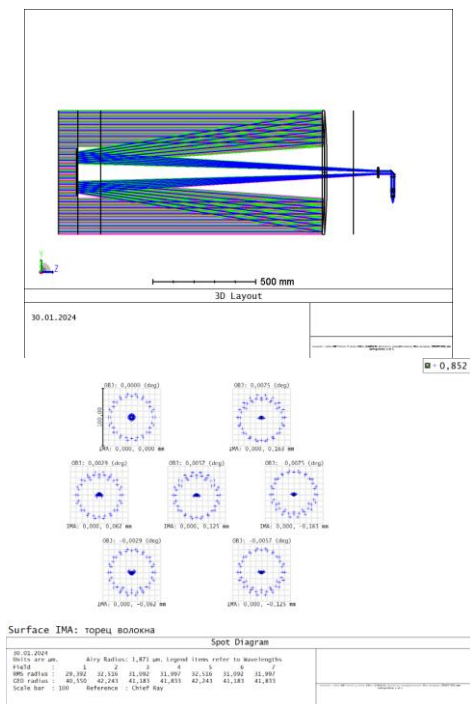


Рис. 1. Оптическая схема наземной станции и форма пятен излучения в фокальной плоскости

Схема состоит из зеркальной оптической системы, афокальной линзы, дихроичного зеркала для спектрального разделения каналов, один из которых отводится на камеру CMOS для работы точного контура наведения, а второй, непосредственно информационный, через спектральный узкополосный фильтр и светоделиитель отбора оптической мощности напрямую попадает на коллиматор оптического волокна, по которому сигнал заводится в передатчик. Светоделиитель нужен для контроля попадания энергии в приемный канал. Схема позволяет получать требуемые размеры пятен излучения в фокальной плоскости по всему полю. Характеристики наземной станции указаны в таблице ниже:

Таблица 2. Основные оптические характеристики наземной станции

| Характеристика | Значение |
|-------------------------|----------|
| Диаметр вх. зрачка, мм | 600 |
| Фокусное расстояние, мм | 1230 |
| Поле зрения, мкрад | 320 |

Физическая реализация схемы и полученные характеристики.

После расчета оптической схемы были изготовлены, закуплены, корпусированы ее элементы и про-

водилась точная настройка схемы (юстировка) в лаборатории квантовых систем НИТУ МИСИС. Измеренные характеристики схемы подтвердили точность ее расчета.

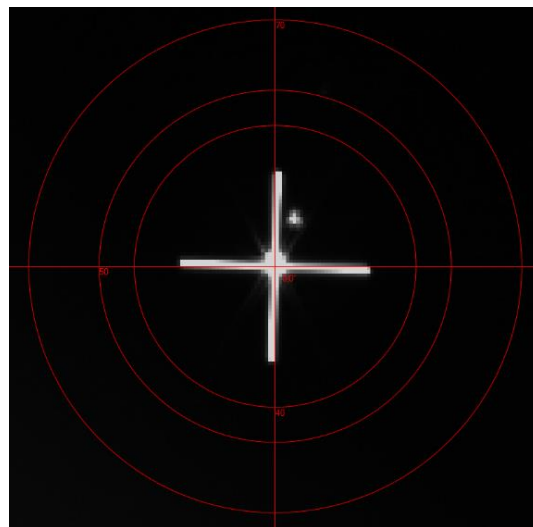


Рис. 2. Процесс юстировки оптической схемы

После юстировки схема была успешно сопряжена с телескопом и готова для работы с мка. Надеемся, при проведении предстоящих будущих сеансов связи с МКА и получения статистически верных данных будут опубликованы результаты работы, связанные с практическим применением системы.

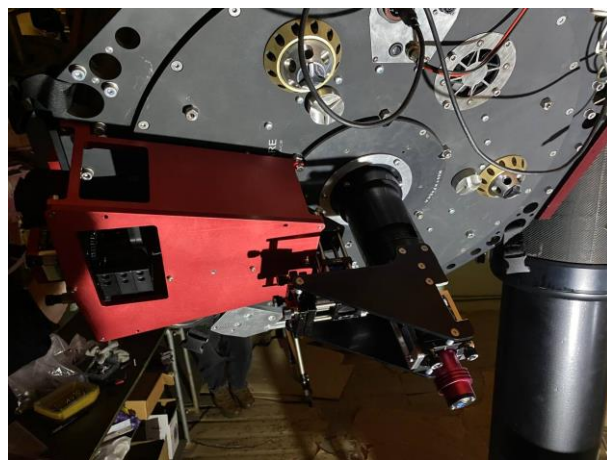


Рис. 3. Установленная схема на телескоп

Заключение

В процессе работы была разработана и изготовлена схема, которая позволяет работать с оптическим сигналом мка Импульс-1 малого формата «кубсат» в том числе для демонстрации квантового канала связи Спутник-Земля для различных условий работы канала. Дополнительно, данная схема позволяет работать с однофотонным сверхпроводящим детектором для регистрации ослабленных до однофотонного состояния оптических импульсов, что открывает большие возможности и перспективы использования.

Список литературы

1. Miller A. V. et al. Vector—towards quantum key distribution with small satellites //EPJ Quantum Technology. – 2023. – Т. 10. – №. 1. – С. 52.
2. Шереметьев А.Г., Толпарев Р.Г. Лазерная связь. – М.: «Связь», 1974.
3. Иванов Н. М., Лысенко Л. Н. Баллистика и навигация космических аппаратов. 2-е изд., перераб. и доп. М. : Дрофа, 2004.
4. Михельсон Н.Н. Оптика астрономических телескопов и методы ее расчета. -М.: Физматлит, 1995.-333с.
5. Joseph M. Geary. Introduction to Lens Design with Practical ZEMAX Examples. Willmann-Bell, 2002. -462с

ТЕСТИРОВАНИЕ АППАРАТНОЙ ЧАСТИ ТРАНСПОРТНОГО РОБОТА С ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМОЙ УПРАВЛЕНИЯ

Желтова Д.Н.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе рассматриваются наиболее распространенные методы тестирования аппаратной части транспортных роботов с интеллектуальной системой управления, такие как визуальный осмотр, внутрисхемное тестирование, тестирование методом "летающих щупов", JTAG-тестирование, функциональный тест и самодиагностика. В качестве рационального метода выбрано JTAG-тестирование совместно с функциональным тестированием, показана важность автоматизации процесса тестирования.

Введение

Транспортные роботы с интеллектуальной системой управления (ИСУ), способные автономно функционировать в неопределённых условиях [1], представляют большой интерес для целого ряда различных практических приложений, главными из которых являются:

- доставка продуктов питания и почты;
- автоматизация производственных процессов;
- ликвидация последствий аварий и катастроф и др.

Модули транспортных роботов с интеллектуальной системой управления

Транспортные роботы с ИСУ представляют собой сложные программно-аппаратные системы, включающие множество модулей. Каждый из модулей требует тщательной проверки, поскольку от его работоспособности зависит функционирование всей системы.

В качестве примера на рис.1 приведена структурная схема современного робота-доставщика [2].

В основе структурной схемы находится вычислительное устройство, осуществляющее руководство

действиями робота посредством алгоритмов управления. В вычислителе обрабатываются видеопотоки, полученные с камер.

Контроллер платформы обеспечивает питание платформы, регулирует ток в каждой ветке питания, переключает питание на резервный источник при необходимости. Кроме того, он осуществляет управление колёсами и сбор информации от ультразвуковых сенсоров (парктроников).

Контроллеры колёс получают от контроллера платформы информацию о требуемой скорости каждого колеса и управляют токами в обмотках, обеспечивая соблюдение заданных параметров скорости при различных условиях движения.

Контроллер периферии регулирует работу мотора крышки, замка и бортового света.

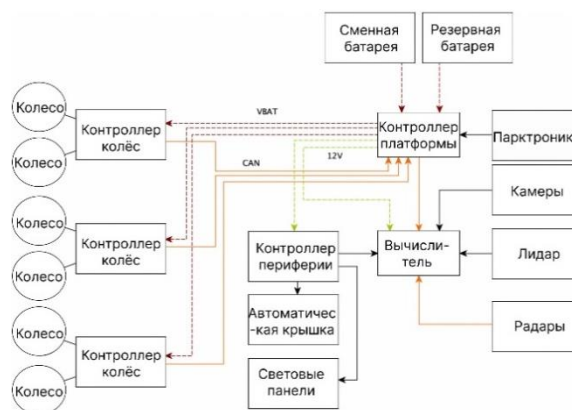


Рис.1. Структурная схема беспилотного робота-доставщика

Классификация методов тестирования

Аппаратная часть транспортного робота с ИСУ может быть проверена различными способами.

1. Визуальный осмотр – это проверка качества монтажа, наличия всех элементов и их целостности. Данный метод позволяет определить недостатки, видимые визуально. Преимуществом данного метода является отсутствие необходимости разработки тестовой оснастки. Однако, высокая вероятность ошибок и низкая производительность снижают эффективность данного метода.

2. Внутрисхемное тестирование. Данный метод использует фиксированные сенсорные щупы для проверки целостности паяных соединений [3]. Достоинствами данного метода являются низкая вероятность ошибок и высокая производительность. К недостаткам метода можно отнести высокую стоимость стенда, большие габаритные размеры, необходимость изготавливать матричное контактное поле в соответствии с конструкцией печатной платы.

3. Тестирование методом "летающих щупов". Хотя данный метод может быть эффективным, так как имеет низкую вероятность ошибок и высокую производительность, его применение ограничено из-за высокой стоимости стенда и необходимости составления кода для перемещения щупов.

4. JTAG-тестирование. Тест соединений через JTAG проверяет наличие непредусмотренных разрывов

вов цепей и лишних замыканий. Тестирование происходит путём выставления на контактах микросхем определённых значений и сравнения считанных значений с заданной таблицей истинности [4]. Данный метод имеет низкую вероятность ошибок, высокую производительность, низкую стоимость, прост в переналадке, так как разработка специализированной оснастки не требуется.

5. Функциональный тест. Является обязательным, так как результаты данного теста свидетельствуют о том, соответствует ли работа электроники установленным требованиям.

6. Самодиагностика является одним из распространенных методов тестирования автомобильной и роботизированной техники. Для построения системы самодиагностики все компоненты системы соединяются с тестовой платой, выполняющей тестирование всех компонентов системы циклично. Основной задачей системы самодиагностики является не поиск причины ошибки, а обнаружение выхода контрольного параметра за пределы допустимых значений с последующим оповещением и кратким отчетом об ошибке. Разработка системы самодиагностики является дорогостоящей, поэтому она применяется в случае большого серийного производства, например автомобилей, но не подходит для транспортных роботов с ИСУ, архитектура которых постоянно изменяется [5].

Таким образом, в случае разработки и производства транспортных роботов с ИСУ, когда имеется много проектов плат, выпускаемых малыми партиями, рационально выбрать метод, который имеет низкую вероятность ошибок, высокую производительность, при этом не требует разработки специальной оснастки, имеет низкую стоимость и прост в переналадке стенда для каждой платы. Всем перечисленным критериям соответствует JTAG-тестирование совместно с функциональными тестами.

Выбранный метод можно автоматизировать, что приведёт к повышению эффективности и скорости проверки печатных узлов. Процесс тестирования будет заключаться в следующем: оператор подключает к тестируемой плате тестовую плату, тестовая плата проводит диагностику, и в случае её успешного завершения тест считается пройденным.

К преимуществам автоматизации тестирования можно отнести:

– Исключение влияния человеческого фактора. При автоматизированном тестировании участие человека в процессе минимизируется, что позволяет производить проверку в одинаковых, контролируемых условиях. Это делает результаты теста более достоверными.

– Уменьшение временных затрат. Повторяющееся тестирование может проводиться в значительно более короткие сроки.

– Повышение прозрачности процесса. Сбор и анализ данных о процессе и результатах тестирования способствует более точному нахождению неисправностей, а также позволяет понять, какие аспекты тестирования требуют дополнительного внимания.

Заключение

В рамках данной работы рассмотрены различные подходы к организации тестирования аппаратной части транспортного робота с ИСУ, а именно визуальный осмотр, внутрисхемное тестирование, тестирование методом "летающих щупов", JTAG-тестирование, функциональный тест и самодиагностика. В качестве рационального метода выбрано JTAG-тестирование совместно с функциональным тестированием.

В контексте современных разработок транспортных роботов с ИСУ автоматизация тестирования представляет собой значительную выгоду, так как существенно повышает качество и эффективность процесса.

Список литературы

1. Экспериментальный образец транспортного мини-робота с интеллектуальной системой управления / Лохин В. М., Евстигнеев Д. В., Манько С. В. [и др.]. - Текст : непосредственный // Интеллектуальные системы автоматического управления / Под ред. И.М. Макарова, В.М. Лохина – М ФИЗМАТЛИТ, 2001 – С. 479-493.

2. Встречаем ровер третьего поколения: история создания робота-курьера Яндексa - Текст: электронный // Хабр : [сайт] – 2021. - 25 ноя - URL:

<https://habr.com/ru/companies/yandex/articles/590997/> (дата обращения 28.01.2024)

3. Соловьев А. Методы тестирования печатных плат и электроники до и во время производства / А. Соловьев - Текст: электронный // КЕДР SOLUTIONS : [сайт] - 2024. - URL: <https://kedrsolutions.ru/blog/metody-testirovaniya-pechatnyh-plat-i-elektroniki#rec581031023> (дата обращения 28.01.2024)

4. Что такое JTAG и как это применить для тестирования печатных плат? – URL: <https://www.xjtag.com/ru/about-jtag/what-is-jtag/> (дата обращения 28.01.2024)

5. Araujo, Hugo & Mousavi, Mohammad & Varshosaz, Mahsa. (2022). Testing, Validation, and Verification of Robotic and Autonomous Systems: A Systematic Review. ACM Transactions on Software Engineering and Methodology. 32. 10.1145/3542945.

АВТОМАТИЗАЦИЯ ОБНОВЛЕНИЯ КАРТ ДОРОЖНОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ GPS-ТРЕКЕРА

*Чайкин Л.Б., Подкопаева П.О., Маршутина Е.Н.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе предлагается решение важной в картографии задачи – актуализация геоданных дорожной сети. В работе предлагается использование программно-аппаратного комплекса с применением

GPS-трекеров. Спроектирована система, позволяющая осуществлять сбор и обработку геоданных двумя способами, а также принимать решение об актуализации данных карты с помощью web-интерфейса.

Введение

В настоящее время всё более актуальной становится проблема использования, обновления и верификации открытых данных. В масштабах страны геоданные устаревают достаточно быстро [1], поэтому важной задачей в данной области является их своевременная актуализация. Заделом в области является волонтерский проект OpenStreetMap [2]. Картографическими данными проекта OpenStreetMap ежедневно пользуются миллионы людей как для профессиональных исследовательских целей, так и для регулярных текущих задач, таких как построение маршрутов или поиск различных объектов на карте.

В настоящее время задача актуализации карт дорожной сети является весьма трудоемкой, если заниматься ей вручную. Поэтому существует потребность в создании и развитии системы автоматизированного сбора и обработки геоданных.

Одни из такого рода систем – веб-ГИС. Они позволяют получать данные со спутника в режиме онлайн и проводить оперативный анализ данных. Одними из самых известных являются NextGIS [3] и CARTO [4], однако они являются коммерческими продуктами и имеют высокий порог входа в плане профессиональной подготовки.

Также в научном сообществе предлагаются различные методы обновления баз данных дорожных сетей. Например, в работах [5, 6] описан подход, позволяющий осуществлять реконструкцию дорожной сети по аэрофотоснимкам благодаря объединению обработки цифровых изображений и информации из баз данных. В работе [7] описывается опыт автоматизации обработки геоданных для платформы Open GIS Consortium [8]. Работа [9] содержит общее описание этапов реализации автоматизации обработки и уточнение пространственных данных.

Более того, уже несколько лет проводятся обширные исследования в области ГИС, сбора и фильтрации геоданных, баз геоданных и управления геоданными, а также построения цифровых карт и геовизуализации [10, 11].

Таким образом, существует потребность в создании доступного инструмента для актуализации геоданных дорожной сети. Для достижения цели данной работы могут использоваться GPS-данные с трекеров или мобильных устройств.

Следующие разделы освещают: метод решения проблемы и текущие результаты, потенциал использования результатов, архитектуру разрабатываемой системы.

Метод предлагаемого решения

Решение поставленной задачи возможно путем автоматизации генерации предложений по исправлению карт.

На основе генерируемых карт с собранными GPS-треками и расхождениями в текущей дорожной сети

специалист может своевременно определить необходимость в актуализации и исправить некорректные геоданные, загрузив сгенерированную подложку с треками в программу для редактирования карт.

Для того, чтобы при приеме геоданных с GPS-трекеров сократить количество точек, уменьшить количество выбросов и избежать сильных потерь точности позиционирования, чаще всего используют различного рода фильтры (точности, скорости, медианный, ускорения, сворачивания точек). В программном комплексе предлагается использование метода Калмана, который на данный момент широко применяется в картографии [12]. Он заключается в том, что с помощью него можно рекурсивно оценить точность (в данном случае – координат) с помощью неполных и зашумленных измерений.

На данный момент реализован прототип мобильного приложения для сбора геоданных и отправки их на сервер. В свою очередь, на сервере происходит сохранение получаемых данных в таблицу. Сделана техническая страница для отображения 100 последних измерений с каждого устройства.

Предлагаемое решение заключается в отображении в web-интерфейсе треков и карты с подсветкой неактуальных данных на основе сравнения текущей карты дорожной сети с собранными данными. С помощью этой карты можно принимать решение о необходимости внесения изменений.

Потенциал использования

Потенциальными пользователями разрабатываемой системы являются волонтеры, готовые предоставлять свои GPS-треки, а также картографы, которые могут принять управленческие решения о необходимости актуализации геоданных.

Использование разрабатываемого комплекса позволяет производить сбор геоданных сразу двумя способами, так что можно выбрать подходящий для тех или иных условий работы. Благодаря Android-приложению можно собирать данные координат и отправлять их в базу данных прямо с мобильного устройства. А во время длительных поездок, например, можно устанавливать устройство сбора GPS-данных в средство передвижения и получать координаты/треки с него, обходясь без мобильного устройства и не расходуя его энергию.

Разрабатываемый программно-аппаратный комплекс позволяет упростить работу при актуализации карт дорожной сети. В отсутствие системы с автоматической генерацией предложений по исправлению карт неактуальные данные могут не исправляться достаточно продолжительное время и вводить в заблуждение пользователей до следующего планового обновления.

Архитектура программного комплекса

Предлагается следующая архитектура системы (рис. 1).



Рис. 1. Архитектура программного комплекса

Архитектура разрабатываемого комплекса состоит из 3 частей: системы сбора геоданных, сервера (backend) и web-интерфейса (frontend). Сбор геоданных может происходить как с помощью мобильного приложения на базе Android, так и при помощи соответствующего устройства, представляющего собой одноплатный компьютер с GPS-датчиком. Данные могут передаваться непосредственно в базу данных, расположенную на сервере, как в виде отдельных GPS-координат, так и в виде агрегированных треков в формате GPX [13].

Благодаря одноплатному компьютеру существует возможность сравнить точность геоданных, получаемых с различных GPS-датчиков, протестировав их работу в одинаковых условиях.

Сервер включает в себя базу данных и прикладную логику, которая отвечает за обработку собранных геоданных. Обработка геоданных в свою очередь состоит из 4 шагов:

- 1) Удаление некорректных данных, возникших из-за сбоя GPS-датчика;
- 2) Агрегация данных (объединение близких точек с учётом погрешности GPS-датчика);
- 3) Генерация трека путем объединения переданных отдельных GPS-координат;
- 4) Сохранение трека в базу данных.

Кроме того, сервер позволяет генерировать подложки с треками по заданной области, которые могут быть использованы картографами для актуализации геоданных с помощью специализированного программного обеспечения, в частности программа JOSM – Java OpenStreetMap Editor для редактирования открытых карт OSM (OpenStreetMap).

Сохраненные треки отображаются в web-интерфейсе на отдельной странице. Также присутствует возможность отображения карты с подсветкой неактуальных данных, которая может помочь в принятии решения о необходимости внесения изменений.

Заключение

Разработана часть программного комплекса, а именно прототип мобильного приложения для сбора геоданных. Реализована передача данных на сервер, где они сохраняются в соответствующую таблицу. Также создана техническая страница для отображения 100 последних измерений с каждого устройства.

Разрабатывается устройство на основе одноплатного компьютера с GPS-трекером для сбора геоданных. Представленное решение смогут использовать как волонтеры, так и профессиональные картографы.

В дальнейшем планируется провести дополнительные исследования, посвященные сравнительному анализу точности геоданных с разных GPS-трекеров, исследованию достаточной частоты обновления геоданных, исследованию фильтров для GPS-данных (для повышения точности позиционирования) для улучшения качества представленного программно-аппаратного комплекса.

Список литературы

1. How Often Does Google Maps Update Satellite Images? / [Электронный ресурс] // nerdtechy.com : [сайт]. — URL: <https://nerdtechy.com/how-often-google-maps-update-satellite-images> (дата обращения: 13.02.2024).
2. OpenStreetMap / [Электронный ресурс] // Openstreetmap : [сайт]. — URL: <https://www.openstreetmap.org> (дата обращения: 15.02.2024).
3. Gramrekha Mobile documentation / [Электронный ресурс] // Gramrekha Mobile GIS App : [сайт]. — URL: <https://nextgis-praxisnfp.readthedocs.io/en/latest/> (дата обращения: 16.02.2024).
4. Location Intelligence & GIS for Cloud Natives | CARTO / [Электронный ресурс] // CARTO : [сайт]. — URL: <https://carto.com/> (дата обращения: 16.02.2024).
5. Chunsun Zhang, Towards an operational system for automated updating of road databases by integration of imagery and geodata // ISPRS Journal of Photogrammetry and Remote Sensing, 2004. - Vol. 58. - p. 166-186.
6. Mena, J.B. State of the art on automatic road extraction for GIS update: A novel classification. // Pattern Recognit. Lettets 24, 2003. – p. 3037–3058.
7. S. Shi and N. Walford. Automated Geoprocessing Mechanism, Processes and Workflow for Seamless Online Integration of Geodata Services and Creating Geoprocessing Services // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2012. - Vol. 5 - No. 6. - p. 1659-1664.
8. About - Open Geospatial Consortium / [Электронный ресурс] // ogc.org : [сайт]. — URL: <https://www.ogc.org/about-ogc/> (дата обращения: 18.02.2024).
9. Simon, M., Asche, H.. Automated Spatial Data Processing and Refining // Lamprecht, AL. (eds) Leveraging Applications of Formal Methods, Verification, and Validation, 2016. - Vol 683.
10. Hardy P. High-quality cartography in a commodity GIS: experiences in development and deployment //Proc. First ICA Symposium for Central and Eastern Europe, 2009. – p. 16-17.
11. Buckley A., Frye C., Battenfield B. An information model for maps: towards cartographic production from GIS databases //22nd ICA Conference Proceedings, A Coruña, Spain, 2005.

12. Ganesh Laveti, G. Sasibhushana Rao, Bharati Bidikar. Modified Kalman Filter for GPS Position Estimation over the Indian Sub Continent // Procedia Computer Science, 2016. - Vol. 87. - p. 198-203.

13. RU:GPX - OpenStreetMap Wiki / [Электронный ресурс] // wiki.openstreetmap.org : [сайт]. — URL: <https://wiki.openstreetmap.org/wiki/RU:GPX> (дата обращения: 16.02.2024).

РАЗРАБОТКА SIP СИСТЕМЫ СЛУЖЕБНОЙ СВЯЗИ

Асташов С.Д., Елисеенко А.М.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Данная работа посвящена созданию системы служебной связи – интеркома.

Целью работы является разработка решения, позволяющего организовать голосовую связь во время видеопроизводства. В ходе разработки будут выполнены задачи по выбору архитектуры интеркома, её реализации на программном уровне. Далее программное решение будет

Разрабатываемая система после внедрения призвана устранить проблему с коммуникацией при создании видеоконтента в Медиацентре МИЭМ.

Введение

В сфере видеозаписей и видеотрансляций часто возникает проблема коммуникации между членами съёмочной команды, располагающимися в аппаратной комнате каждый на своём месте, а также между режиссёром, членами съёмочной группы и гостями студии, каждая из которых могут находиться в разных помещениях. Причиной является важная особенность такого продакшена — распределённая сетевая съёмка. Одно из самых простых и действенных решений данной проблемы — интерком. Его внедрение в работу телестудии при грамотном подходе не требует крупных финансовых и технических вложений и позволяет достичь поставленной цели.

Благодаря повсеместному наличию в МИЭМ как проводной, так и беспроводной сети при разработке системы связи есть возможность не рассматривать системы, функционирующие на основе аналоговых сигналов, а сразу перейти к системам, работающим на прикладном уровне.

Так, особенностью данного решения является использование протокола SIP [2] в основе своей работы, что позволяет осуществлять соединение между пользователями поверх сети интернет. Также одно из достоинств разработанного решения — независимость от наличия графического сервера на устройстве, используемом в качестве клиента интеркома, что даёт разработчику возможность внедрить систему связи как на обычные аппаратные устрой-

ства с дисплеем, так и на серверные, у которых видеовыход отсутствует.

Обзор источников

В настоящее время наиболее надёжные архитектуры систем связи базируются на использовании локальных сетей и протокола SIP. Примером тому может послужить задействование таких систем в центрах управления космическим пространством [3]. Также к этим системам прибегают, если организации требуется широкоэмитательный формат коммуникации [4]. Примечательно, что данные системы при развёртывании в пределах здания могут совмещать низкую стоимость, простоту архитектуры и высокую стабильность работы [5]. Эти свойства не противоречат требованиям к разрабатываемой системе, а именно: дешевизне, гибкости, работе на прикладном уровне и возможность интеграции с сервисами МИЭМ.

Архитектура системы

На данный момент подавляющее большинство простых и не только систем передачи голосовой информации работают через SIP. Этот протокол отличается гибкостью и возможностью масштабирования, он обеспечивает создание, модификацию и завершение сеансов между двумя и более участниками. Он работает на прикладном уровне и это даёт возможность использовать его поверх существующих сетей.

Архитектурно любая подобная система состоит из серверной части и клиентов. В качестве SIP-сервера был выбран Asterisk [6]. В настоящее время Asterisk — самый популярный программный SIP-сервер, он бесплатный, поддерживает операционную систему GNU/Linux [7] и занимает 85% рынка подобных серверов. От сервера ожидается, что он будет обеспечивать обслуживание до тридцати абонентов, поддерживать как личные звонки с одного клиентского устройства на другое, так и групповые звонки. Установка сервера планируется на компьютер с архитектурой x86 [8] под управлением операционной системы Linux. Схема взаимодействия клиентов и сервера изображена на рис.1.

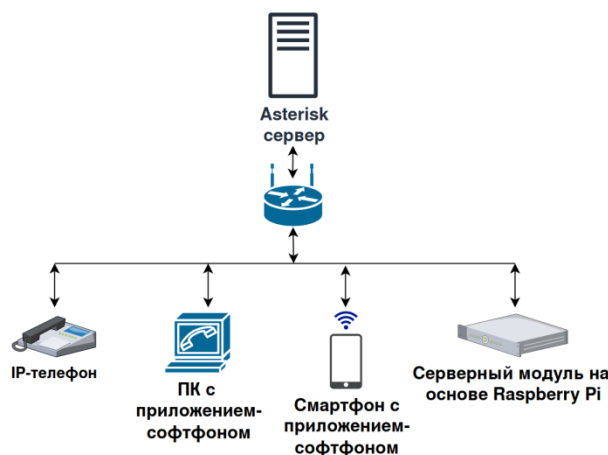


Рис.1. Общая архитектурная схема

В условиях работы съёмочной группы важно не отвлекаться на посторонние действия, работа с приложением-клиентом системы связи должна быть простой, понятной и удобной. В то же время нужно учесть, что для работы на серверных ячейках нужна программа без графического интерфейса, которой можно будет управлять с помощью устройств ввода посредством команд оболочки операционной системы. Итого для обычных устройств нужно кроссплатформенное приложение-софтфон, а для серверных ячеек — headless [9] (без графического интерфейса) клиент с возможностью работы без графики. Схема работы системы в случае с headless-клиентом изображена на рис.2.

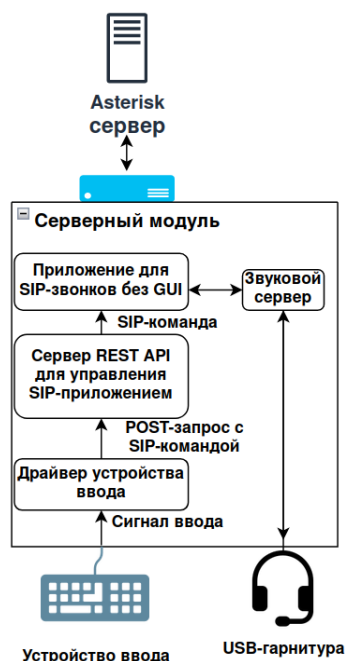


Рис.2. Схема headless-клиента

Нагрузочное тестирование

Тестирование работы SIP системы под нагрузкой проводилось с использованием клиентской утилиты sipp [10] и встроенными возможностями сервера Asterisk. Параметры сервера: ОС – Ubuntu 22.04, CPU – Intel Core i5-9400T. Тип тестирования - генерация одновременных вызовов с длительностью в 100 секунд. При установке соединения сервер, проигрывая аудиофайл отправляет трафик на удалённую сторону, затем те же пакеты перенаправляются обратно. Мониторинг на клиентской части проводился при помощи программ htop [11], nload [12], sipp. Задачей тестирования являлось определение максимального количества одновременных вызовов, при котором система работает стабильно с удовлетворительным качеством звука. В результате максимальное количество одновременных вызовов со стабильным битрейтом в 13 кбит/с – 800 штук. Нагрузка на CPU не превышала 80%. При количестве вызовов >850 в логах сервера появляются критические ошибки.

В сравнении с готовыми проводными решениями, где в комплекте обычно предусматривается до 10 устройств, система на базе Asterisk может обслуживать гораздо большее количество клиентов. А сами клиенты могут быть совершенно разными устройствами. Это даёт преимущество в финансовых затратах на систему связи, а также позволяет интегрировать в интерком сторонние сервисы.

Заключение

В ходе разработки была создана и протестирована система служебной связи в специфике сетевой телестудии. Программный продукт был интегрирован в инфраструктуру Медиацентра МИЭМ.

Список литературы

1. Arm CPU Architecture - Arm® <https://www.arm.com/architecture> (дата обращения: 10.02.2024).
2. RFC 3261: SIP: Session Initiation Protocol <https://www.rfc-editor.org/rfc/rfc3261.html> (дата обращения: 10.02.2024).
3. Keller, C., Bellenberg, U., & Dallat, J. (2012). Incremental deployment of a voice intercom system in a modern space control center. SpaceOps 2012 Conference. <https://doi.org/10.2514/6.2012-1294113>
4. Imazeki, Y., Arai, T., & Narita, N. (2006). A communication system using IP technology for broadcasting stations. In *Kyokai Joho Imeiji Zasshi/Journal of the Institute of Image Information and Television Engineers* (Vol. 60, Issue 2). <https://doi.org/10.3169/itej.60.175>
5. Bai, Y. W., & Chen, S. C. (2007). Design and implementation of home intercom and security control systems for buildings. Proceedings of the International Symposium on Consumer Electronics, ISCE. <https://doi.org/10.1109/ISCE.2007.4382155>
6. Home - Asterisk Documentation <https://docs.asterisk.org/> (дата обращения: 10.02.2024).
7. The GNU Operating System and the Free Software Movement <https://www.gnu.org/> (дата обращения: 10.02.2024).
8. x86 assembly/X86 Architecture https://en.wikibooks.org/wiki/X86_Assembly/X86_Architecture (дата обращения: 10.02.2024).
9. Understanding Headless Architecture: What is it? - Medium <https://medium.com/strapi/understanding-headless-architecture-what-is-it-f02a0aaf999e> (дата обращения: 10.02.2024).
10. GitHub - SIPp/sipp: The SIPp testing tool <https://github.com/SIPp/sipp> (дата обращения: 10.02.2024).
11. htop - an interactive process viewer <https://htop.dev/> (дата обращения: 10.02.2024).
12. GitHub - rolandriegel/nload: Real-time network traffic monitor <https://github.com/rolandriegel/nload> (дата обращения: 10.02.2024).

ОБЗОР МЕТОДОВ РАЗВЕРТЫВАНИЯ СЕТЕЙ ИМЕНОВАННЫХ ДАННЫХ

Сафронов А.М., Михайлов К.В.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе рассматриваются варианты и процесс развертывания топологии, использующей сеть именованных данных (named data network, далее NDN). Данный обзор проводится в рамках исследования работы сетей на базе NDN.

Введение

В настоящее время в интернете находится масса информации. С развитием интернет-магазинов, поисковых систем, файлообменников, стриминговых платформ и прочих вариаций передачи этой информации, существенная часть которой постоянно повторяется и множится, все больше и больше устаревает общепринятый стек протоколов TCP/IP, созданный для передачи информации из одной точки в другую. Из-за повышенного спроса на эффективное распространение информации появился новый подход к передаче контента, появление новых архитектур, основанных на именованных объектах данных (named data object, NDO), что, в свою очередь, повлекло изменение в существующих сетях. Новый подход, известный как информационно-ориентированная сеть (information centric networking, ICN), должен отвечать современным требованиям и призван устранить существующие недостатки привычных TCP/IP [1].

Существует несколько вариантов сетей, являющихся ICN, однако две наиболее популярные и развитые - контентно-ориентированная сеть (content-centric networking, CCN) и сеть именованных данных (named data networking, NDN). Принцип работы у них схожий, однако NDN является идейным последователем CCN, имея в своем арсенале улучшенные алгоритмы кэширования, более оптимизированную маршрутизацию и полностью рабочий прототип.

В данной работе рассматриваются плюсы и минусы протокола NDN, а также варианты создания сетей, основывающихся на данном протоколе.

Анализ сетей именованных данных

Как уже говорилось ранее, использование NDN смещает фокус с IP-адресов на имена данных, соответственно блоки данных запрашиваются по их идентификатору, а не IP-адресу. Также NDN поддерживает кэширование на уровне сети, что позволяет уменьшить нагрузку на всю сеть при повторных запросах. К тому же пакет данных включает в себя такие механизмы безопасности, как, например, цифровая подпись, что повышает безопасность передаваемых данных.

Однако, у NDN есть и проблемы, решить которые пока сложно в силу того, что эта архитектура еще

исследуется и развивается. Наличие механизмов защиты может привести к потере производительности, особенно в том случае, если будет передаваться много небольших пакетов с данными, каждый из которых обладает собственной подписью, поэтому к этому надо подходить аккуратно. Подобное справедливо и по отношению к кэшу. Он снижает нагрузку на сеть, но без правильных инструкций это может привести к проблеме согласования данных между разными узлами.

Ранее перечисленные проблемы, несомненно, являются важными, но при грамотном подходе их можно решить аналитическим способом, не прибегая к перебору различных параметров на практике. В то же время есть еще одна очень важная проблема, решить которую невозможно без практики – проблема масштабирования [2]. Создать реальную большую сеть со множеством узлов, не используя NDN как надстройку над TCP/IP очень тяжело. Помимо того, что нужно найти подходящее количество материальных клиентов в количестве нескольких десятков тысяч, если не сотен, нужно изменить уже существующие прикладные приложения, чтобы они могли эффективно работать в соответствии с новой архитектурой. Данное исследование можно провести и с использованием виртуальных машин, а не реальных, но продолжительное тестирование реальной сети должно дать больше информации, нежели симуляция. В рамках данной статьи невозможно проделать такой объем работы, поэтому ограничимся развертыванием сети на примере небольшой топологии.

Также стоит отметить, что существующие статьи рассматривают сети, в которых не используются какие-либо реальные протоколы, а передача данных идет исключительно посредством NDN. Такое можно увидеть в статье, в которой изучаются политики кэширования ICN сетей [3]. Но есть и статьи, рассматривающие работу NDN и TCP/IP в симбиозе [4].

Развертывание на физическом устройстве

Как уже упоминалось ранее, существует рабочий прототип NDN, с помощью которого можно установить и протестировать работоспособность данной сети. Программное обеспечение находится в открытом доступе и называется Named Data Networking Forwarding Daemon (NFD). Найти его можно на гитхабе, вместе с инструкцией по установке вместе со всеми зависимостями.

После установки с помощью команды “nfd start” необходимо запустить NFD на каждой из машин в рассматриваемой топологии, а также прописать на потребителях и маршрутизаторах интерфейсы, на которые будет идти пакет интереса. После того, как все необходимые интерфейсы будут прописаны, можно проверять работоспособность всей сети с помощью известной команды “ping”, либо же начать проводить собственные исследования, передавая уже полноценные пакеты с информацией или даже чанки информации, содержащиеся в нескольких пакетах сразу. Стоит отметить, что в данном прототипе производитель является постоянно активным сервером, который готов выдавать необходимую информацию,

в то время как потребитель отправляет пакет интереса по команде, после чего ему приходит ответ. Также на любой топологии, использующей одно из устройств в качестве маршрутизатора, можно увидеть, что пакет с запрашиваемой информацией приходит к потребителю без заранее прописанного пути, используя лишь интерфейсы, по которым прошел пакет интереса.

Еще одна важная деталь, про которую стоит упомянуть – возможность настроить политику кэширования с помощью команды “nfdc cs”. Среди параметров можно найти максимальное количество записей в кэше, включение и отключение возможности делать новые записи в кэше, а также опцию ответа на пакет интереса с помощью данных, находящихся в кэше. Помимо переключаемых параметров можно найти количество текущих записей в кэше, а также количество промахов и попаданий, дающее возможность измерять различные политики кэширования. Для примера, рассмотрим сравнение работоспособности линейной сети со включенным кэшированием и без (см. рис.1).

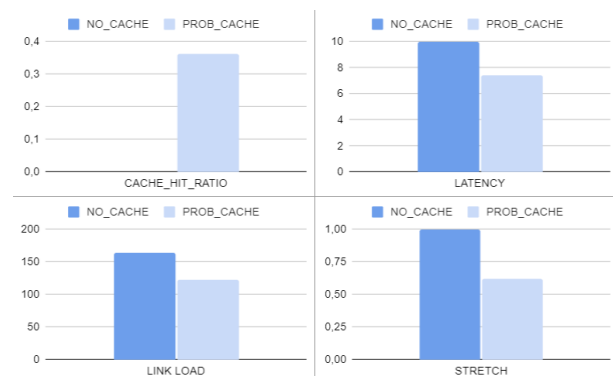


Рис.1. Сравнение двух политик кэширования

Развертывание на виртуальной машине

Существует несколько вариантов собрать сеть, основанную на NDN, на виртуальной машине. Mini-NDN представляет собой инструмент сетевой эмуляции, который позволяет тестировать и исследовать платформу NDN, основанную на Mininet. По сути своей Mini-NDN использует все те же библиотеки, что были описаны ранее. Главное его преимущество – отсутствие необходимости в нескольких устройствах, ведь вся сеть эмулируется на одном устройстве, а ее содержимое в виде количества устройств, политики кэширования, интерфейсов легко настраивается отдельно. Стоит также отметить, что установить Mini-NDN даже проще, чем NFD, ведь разработчики этой утилиты упростили установку всех зависимостей до одной команды.

Заключение

Хоть NDN и находится в ранней стадии исследования, у него уже есть работоспособный прототип, позволяющий исследовать работу сети на его основе. Однако, все еще большой проблемой является его распространение и глобальное использование, ведь

прототип не рассчитан на большое количество одновременно подключенных устройств.

Список литературы

1. Боровская Я. А., Гребешков А. Ю. ICN и сети именованных данных //V Научный форум телекоммуникации: теория и технологии ТТТ-2021. – 2021. – С. 103-104.
2. Равиндран Р. и др. Способ и устройство для масштабируемой маршрутизации контента и мобильности в сетях именованных данных. – 2018.
3. Навроцкий Я. Ю., Пацей Н. В. Реализация политик кэширования в информационно-ориентированных сетях //Труды БГТУ. Серия 3: Физико-математические науки и информатика. – 2018. – №. 1 (206). – С. 99-103.
4. Блинников М. А., Пирмагомедов Р. Я. Сетевой шлюз для сетей именованных данных //Электросвязь. – 2021. – №. 5. – С. 22-30.

РЕКОНФИГУРИРУЕМЫЕ СТЕНДЫ

*Горовой В.В., Прокудина К.В.,
Апыюк В.Р., Огай В.А.*

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Исследована эффективность различных методов и алгоритмов дистанционной работы с оборудованием. Создание и внедрение реконфигурируемого стенда в учебный процесс является одним из важных частей для улучшения качества обучения студентов в настоящее время. Для того, чтобы данное оборудование было доступно было принято решение разработать реконфигурируемые стенды для синхронной/асинхронной работы с оборудованием.

Введение

В современном мире все чаще возникают ситуации, когда студент или сотрудник нуждается в удаленном доступе к лабораторным стендам. Причинами возникновения такой необходимости является, как долгая дорога к лаборатории, так и распространение различных вирусных инфекций. Эти причины и сподвигли сделать аппаратно-программный комплекс, который нацелен на выполнение работ в рамках лабораторной и проектной деятельности, а также задачи, описанные в книге доцента НИУ ВШЭ Романова А.Ю. «Цифровой синтез» [1]. Комплекс (рис.1) можно модернизировать под другие виды практических заданий, требующих для выполнения специфическое программное обеспечение и плату. Создание серверной стойки с ячейками для удаленной работы с платами ПЛИС [2] позволит эффективно поддерживать учебный процесс не только в УЛ САПР МИЭМ НИУ ВШЭ [3], но и в других местах, где сталкиваются с такой проблемой, а также сэкономит трудо-, энергозатраты и материальные ресурсы.

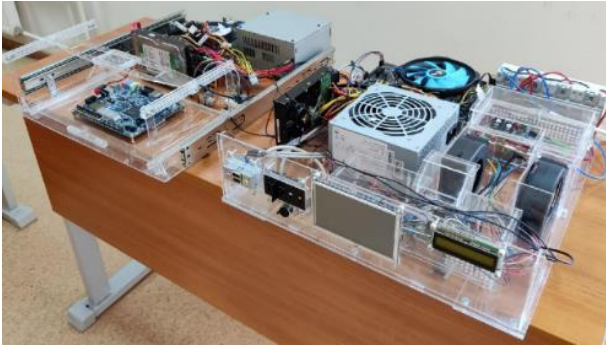


Рис.1. Собранные второстепенная (слева) и главная (справа) ячейки

Выявление проблемы

Важным этапом разработки удаленной лаборатории является асинхронность и создание грейдера, что позволяет обеспечить более эффективную работу и ускорить процесс проверки работ студентов. Стенд работает в 2 режимах: синхронном и асинхронном. Асинхронность [4] позволяет выполнять несколько задач одновременно, что увеличивает производительность системы и сокращает время ожидания. Грейдер автоматически проверяет работы студентов на соответствие заданным критериям и собирает статистику в удобный для преподавателя вид, что также ускоряет процесс проверки и облегчает проверку работ. Важно обеспечить возможность обратной связи со студентами, чтобы они могли получить дополнительные рекомендации и советы от преподавателя по улучшению своей работы.

Обзор литературы

С начала пандемии коронавируса наблюдается рост популярности удаленных лабораторий на базе ПЛИС. Это связано с необходимостью обеспечения безопасности и удобства для исследователей и студентов, которые не могут посещать физические лаборатории из-за ограничений, связанных с пандемией. Удаленные лаборатории на базе ПЛИС позволяют проводить эксперименты и измерения, управляя оборудованием через Интернет. Это открывает новые возможности для обучения и исследований.

Академическая работа по данному вопросу приобретает все большее значение. Ученые и преподаватели активно ищут способы интеграции удаленных лабораторий на базе ПЛИС в образовательные программы и исследовательские проекты. Это требует разработки новых методов обучения, адаптации программного обеспечения и создания инфраструктуры для удаленного доступа к лабораторным помещениям.

Создание и внедрение удаленной лаборатории с использованием технологии ПЛИС может улучшить обучение цифровой электронике. Это предоставляет студентам веб-платформу для удаленной оценки проектов с использованием реального оборудования. Выделенный сервер [5], подключенный к системе разработки, настраивает программируемое цифровое устройство и позволяет удаленным пользователям взаимодействовать с периферийными устройствами. Сервер также управляет подключениями и их про-

должительностью, предотвращая перегрузку ресурсов и облегчая сбор статистики использования. Это достижение в области образовательных технологий предоставляет возможность улучшить процесс обучения в области цифровой электроники. Есть веб-доступ к стенду, но стремимся предложить еще более динамичную и интерактивную среду обучения, внедряя возможность реконфигурации в платформу удаленной лаборатории.

Система, описанная в работе [6], позволяет студентам удаленно получать доступ к лабораторному оборудованию через Интернет с помощью VPN-соединения. Каждая плата FPGA DE 10-Lite подключена к плате Arduino Due для доступа к сигналам ввода-вывода. Студенты используют программное обеспечение Quartus Lite для разработки проектов на плате FPGA. Приложение Matlab отправляет управляющие сигналы на плату Arduino Due, чтобы продемонстрировать функциональность FPGA DE 10-Lite с помощью веб-камеры. Это позволяет студентам участвовать в практическом обучении, несмотря на физические ограничения. Наша система также включает использование веб-камеры в демонстрационных целях. Она состоит из двух основных компонентов: платы FPGA и Arduino. Однако в нашей системе плата FPGA служит основным модулем, в отличие от платы Arduino, описанной в статье.

В работе [7] представлена альтернатива для удаленных лабораторий FPGA под названием Pitanga. Pitanga использует удаленный эмулятор FPGA, который работает на стандартных процессорах x86 вместо физической платы FPGA. Это программное решение обеспечивает большую масштабируемость и экономическую эффективность по сравнению с аппаратными системами.

В литературе было предложено множество сопоставимых методологий [8-10]. Эти подходы представляют аналогичную концепцию для внедрения удаленных лабораторий. В работе [11] каждый стенд оснащен сервером для программирования платы FPGA через USB-Blaster и Raspberry Pi для управления переключателями на платах Cyclone IV и Cyclone FPGA. Эти станции также оснащены устройством захвата изображений [12] для применения алгоритмов обработки изображений, а также амперметром для измерения тока в сердечнике через интерфейс RS232. В отличие от использования ПЛИС, Angrisani и др. [13] представили исследование по внедрению гибкого удаленного лабораторного комплекса для обучения в области измерений и приборостроения. Этот комплекс использует программируемую аналоговую матрицу (ПЛАМ) для облегчения модификации аналоговых схем и проведения удаленных экспериментов.

iLabs [14] – это проект, который представляет собой удаленные лаборатории, обеспечивающие доступ к реальному лабораторному оборудованию через Интернет. Они созданы с целью расширения возможностей научного и инженерного образования путем предоставления студентам доступа к разнообразным лабораторным экспериментам, которые ра-

нее могли быть ограничены географическими или финансовыми причинами.

Проект iLabs предоставляет программные инструменты для проведения онлайн-лабораторных экспериментов. Архитектура iLabs состоит из трех модулей: лабораторного сервера, лабораторного клиента и сервисного брокера. Лабораторный сервер управляет лабораторным оборудованием и обеспечивает доступ к нему через интернет. Лабораторный клиент предоставляет интерфейс для работы с лабораторией и позволяет студентам проводить эксперименты и анализировать результаты. Сервисный брокер обеспечивает обмен данными между клиентом и сервером, а также предоставляет общие услуги хранения и администрирования. Проект iLabs был завершен в 2019 году.

VirtualBench от National Instruments [1515] - это универсальное устройство, объединяющее несколько измерительных и тестовых приборов в одном корпусе. Оно предоставляет возможности осциллографа, мультиметра, источника питания, функционального генератора сигналов и цифрового ввода/вывода. VirtualBench упрощает и ускоряет процесс измерений и тестирования, предоставляя широкие возможности для работы с электроникой и автоматизации измерений. Архитектура VirtualBench включает аппаратную часть - измерительные и тестовые приборы, объединенные в единое устройство, и программное обеспечение, обеспечивающее управление и визуализацию данных. Это позволяет пользователям получать доступ к различным функциям измерения и тестирования через удобный и интуитивно понятный интерфейс.

В приведенном обзоре литературы показано, что удаленные лаборатории на базе ПЛИС имеют потенциал для расширения опыта обучения в области цифровой электроники. Также были внедрены альтернативные подходы, такие как использование виртуальных эмуляторов ПЛИС и программируемых аналоговых матриц (ПЛИМ), предлагающие экономически эффективные и масштабируемые решения для удаленного доступа к лабораторному оборудованию. В целом, достижения в области удаленных лабораторий на базе ПЛИС открывают многообещающие возможности для интерактивного и практического обучения в различных областях, включая цифровую электронику, измерения и приборостроение.

Расширение задач

Дальнейшее развитие проекта может включать в себя такие улучшения, как работа с функциональностью серверной стойки, добавление поддержки новых типов плат ПЛИС, расширение возможностей программного обеспечения для работы с платами, создание новых возможностей для удаленной работы, оптимизация работы сетевых протоколов и улучшение аппаратной части шкафа. Также желательно проводить регулярное обновление программного обеспечения и производить тестирование системы на ошибки и уязвимости.

Обновление системы менеджмента питания

Стоит отметить, в серверной стойке помимо четырех второстепенных ячеек есть одна главная, которая контролирует работу всех остальных и в случае аварийных ситуаций отключит питание на них с помощью разработанной системы менеджмента питания (рис.2), представляющей собой «умные розетки» [17]. В эту систему был внедрен датчик напряжения для более точного показания данных. Также рассматривается возможность усовершенствования устройства: Arduino Nano заменить на модуль ESP32 [18], у которого существует Bluetooth, Wi-Fi, еще стоимость его в два раза ниже, чем у первого микроконтроллера. Кроме того, розетку можно тоже сделать универсальной, если добавить в нее промышленный протокол ModBus [19], служащий для организации связи между электронными устройствами и работающая везде за счет своей многоплатформенности и гибкости.

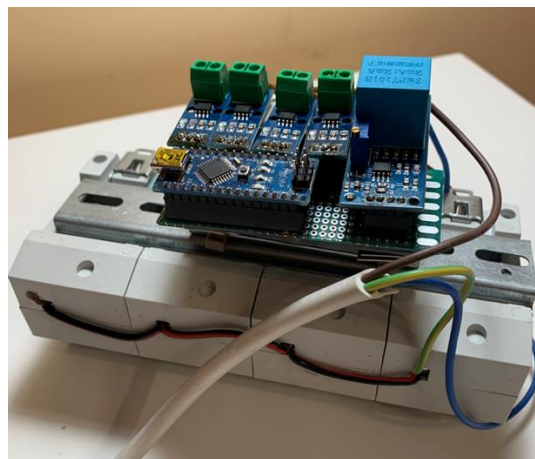


Рис.2. Итоговая сборка систем менеджмента питания

Заключение

В ходе исследования был проведен обзор существующих решений в области создания серверной ячейки для удаленной работы с платами ПЛИС. Выводы подчеркивают потенциальный успех проекта при наличии совместных усилий команды и выполнении поставленных задач по модернизации, что позволит улучшить процесс выполнения студентами практических заданий и проверки работ преподавателями.

Список литературы

1. Цифровой синтез. Практический курс. – URL: <https://publications.hse.ru/mirror/pubs/share/direct/430635780.pdf> (дата обращения: 15.06.2023).
2. Схема > Плата > ПЛИС. – URL: <https://habr.com/ru/articles/242531/> (дата обращения: 15.06.2023).
3. Проект №859. – URL: <https://cabinet.miem.hse.ru/#/project/859/> (дата обращения: 15.06.2023).
4. Асинхронность как основной подход к разработке высоконагруженных приложений. – URL:

<https://habr.com/ru/articles/733154/> (дата обращения: 15.06.2023).

5. Ó. Oballe-Peinado, J. Castellanos-Ramos, J. A. Sánchez-Durán, R. Navas-González, A. Daza-Márquez and J. A. Botín-Córdoba, "FPGA-Based Remote Laboratory for Digital Electronics," 2020 XIV Technologies Applied to Electronics Teaching Conference (TAE), Porto, Portugal, 2020, pp. 1-5, doi: 10.1109/TAE46915.2020.9163676.

6. A. Flores, M. Raffo, M. Balcazar and K. Ylla-huamán, "Remote Laboratory for Teaching Digital Design using a VPN and embedded system," 2021 IEEE XXVIII International Conference on Electronics, Electrical Engineering and Computing (INTERCON), Lima, Peru, 2021, pp. 1-4, doi: 10.1109/INTERCON52678.2021.9532871.

7. A. S. Costa, L. D. Silveira and A. I. Reis, "Live Demonstration: Pitanga Platform for Virtual FPGA Remote Laboratories," 2023 IEEE International Symposium on Circuits and Systems (ISCAS), Monterey, CA, USA, 2023, pp. 1-5, doi: 10.1109/ISCAS46773.2023.10181345.

8. M. Reichenbach, M. Schmidt, B. Pfundt and D. Fey, "A New Virtual Hardware Laboratory for Remote FPGA Experiments on Real Hardware", Proceedings of the 2011 International Conference on E-learning E-business Enterprise Information System E-Government (EEE), pp. 17-23, 2011.

9. H. H. Saliah, M. Saad, H. Hassan and I. de la Teja, "Virtually and remotely accessing and controlling laboratory real devices: a new trend in teaching and learning in engineering", Proceedings of the International Conference on Information Technology Based Higher Education and Training, pp. 107-114, 2000.

10. J. García Zubía, "Educational software for digital electronics: BOOLE-DEUSTO", Proceedings - 2003 IEEE International Conference on Microelectronic Systems Education: Educating Tomorrow's Microsystems Designers. MSE 2003, pp. 20-22, 2003.

11. A. Schwandt and M. Winzker, "Make it Open - Improving Usability and Availability of an FPGA Remote Lab," 2019 IEEE Global Engineering Education Conference (EDUCON), Dubai, United Arab Emirates, 2019, pp. 232-236, doi: 10.1109/EDUCON.2019.8725098.

12. C. A. Mayoz, A. L. da Silva Beraldo, A. Villar-Martinez, L. Rodriguez-Gil, W. F. M. de Souza Seron and P. Orduña, "FPGA remote laboratory: experience of a shared laboratory between UPNA and UNIFESP", 2020 XIV Technologies Applied to Electronics Teaching Conference (T ABE), pp. 1-8, 2020.

13. Leopoldo Angrisani, Francesco Bonavolontà, Mauro D'Arco, Annalisa Liccardo, "A flexible remote laboratory with programmable device under test", Measurement, Volume 156, 2020, 107584, ISSN 0263-2241, <https://doi.org/10.1016/j.measurement.2020.107584>.

14. iLab Project at MIT. – URL: <https://icampus.mit.edu/projects/ilabs/> (дата обращения: 15.02.2024)

15. Emerson: what is VirtualBench – URL: <https://www.ni.com/en/shop/electronic-test-instrumentation/virtualbench/what-is-virtualbench.html> (дата обращения: 15.02.2024)

16. Alt Linux Wiki. – URL: https://www.altlinux.org/Главная_страница (дата обращения: 15.06.2023).

17. Умная розетка: что можно подключить и как автоматизировать свой дом. – URL: <https://club.dns-shop.ru/blog/t-238-drugoe-dlya-umnogo-doma/53612-umnaya-rozetka-chto-mojno-podkluchit-i-kak-avtomatizirovat-svoi/> (дата обращения: 15.06.2023).

18. ESP32. – URL: <https://ru.wikipedia.org/wiki/ESP32> (дата обращения: 15.06.2023).

19. ModBus. - URL: <https://ru.wikipedia.org/wiki/Modbus> (дата обращения: 15.06.2023).

РАЗРАБОТКА СИСТЕМЫ ВИРТУАЛЬНОЙ СТУДИИ ДЛЯ ВИДЕОПРОИЗВОДСТВА В РЕАЛЬНОМ ВРЕМЕНИ

Айрапетян Т.Ж.¹, Пардаев А.А.¹, Шубекова А.Б.²
Национальный исследовательский университет
«Высшая школа экономики»,
¹департамент компьютерной инженерии,
²департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В данной работе исследуются возможности развития технологии видеопроизводства с использованием виртуальной студии, развернутой на движке Unreal Engine 5, PTZ-камер и хромакея. Практическая разработка подразумевает под собой создание приложения для реализации виртуального 3D фона в условиях съемки в студии в режиме реального времени.

Введение

В настоящее время виртуальное видеопроизводство [1] занимает все более важную роль в создании цифровых видеоматериалов. Требования к высокому качеству и креативности контента подталкивают к разработке новых технологических решений. Одной из таких тенденций является использование виртуальных студий для создания видеоматериалов в режиме реального времени без необходимости последующего монтажа видеоматериалов.

Цель настоящей работы заключается в исследовании и применении возможностей виртуальной студии в сфере видеопроизводства. Основная задача работы - разработать систему, основанную на движке Unreal Engine 5 [2], PTZ-камерах [3] и использовании студии с хромакеем [4], что позволит создавать и использовать виртуальные 3D фоны в реальном времени при съемке в студии.

Главным принципом данной системы является использование виртуальных камер для создания трехмерной сцены и автоматизированные механизмы для подготовки и наложения виртуального фона. Это позволит производить трансляции в реальном времени, снижая необходимость в последующей обработке и обеспечивая высокое качество и реалистичность контента. Использование хромакея в сочетании с технологиями Unreal Engine 5 позволяет создавать реалистичные виртуальные сцены и фоны, которые могут адаптироваться к любым условиям съемки.

Результаты этой работы будут полезны как для профессионалов в области видеопроизводства, так и для широкой аудитории, заинтересованной в создании высококачественного анимированного 3D видеоконтента в режиме реального времени.

Аналоги

В ходе работы было рассмотрено несколько аналогов и альтернативных решений. Один из них - "Anti-latency cam tracker" [5] - основан на использовании высокоскоростных сенсоров, которые непрерывно отслеживают движение камеры. Эти сенсоры могут быть интегрированы непосредственно в камеру или установлены отдельно и связаны с камерой посредством специального интерфейса.

Также было рассмотрено устройство Vive Camera Tracker Sony [6]. Оно использует оптическую систему для отслеживания движения камеры и оснащено камерой высокого разрешения и датчиками глубины. Данные о движении камеры передаются на компьютер или игровую консоль, которые обрабатывают информацию и синхронизируют положение камеры с виртуальным окружением. Это позволяет пользователям свободно перемещаться в виртуальном пространстве и взаимодействовать с объектами, сохраняя при этом точную траекторию движения камеры.

Преимущества и недостатки аналогов

Обзор аналогов наглядно показывает, что на рынке преобладают физические устройства для трекинга, монтируемые напрямую в камеру. Недостаток таких аналогов для применения в рамках данной работы заключается в высокой стоимости и в необходимости докупать дополнительный комплект оборудования для каждой камеры, используемой в виртуальной студии. Также обзор демонстрирует, что на рынке отсутствуют конкурентоспособные программные реализации, обеспечивающие минимальную временную задержку трекинга, минимальную стоимость и высокий показатель качества изображения.

Предлагаемая разработка

Для создания оптимального способа решения данной проблемы были учтены следующие необходимые факторы:

1. Обеспечение минимальной временной задержки трекинга камеры (до 100 мс) и качество съемки, сопоставимое со студийным;
2. Минимальные финансовые затраты на реализацию технологии;
3. Максимальное количество подключаемых одновременно камер;
4. Независимая работа приложения и его масштабируемость. Приложение должно быть самостоятельным и предоставлять возможности разворачивания на внутренних серверах МИЭМ;
5. Простота управления приложением и создания новых 3D сцен.

Для разработки виртуальных сцен была выбрана игровая среда Unreal Engine 5, представляющая собой современный кроссплатформенный движок, созданный компанией Epic Games. Этот выбор обу-

словлен богатым функционалом базовой версии программы, который предоставляет широкие возможности для создания игр и приложений:

1. Реалистичные 3D сцены. Unreal Engine 5 позволяет создавать высоко детализированные и реалистичные трехмерные сцены с помощью передовых технологий визуализации;

2. Поддержка виртуальных камер. Движок предоставляет возможность работы с виртуальными камерами, что позволяет операторам реализовывать динамичные кадры и взаимодействовать с виртуальным пространством в реальном времени;

3. Широкие возможности настройки освещения. С помощью системы динамического освещения и инструментов настройки освещения можно достичь высокой степени реализма в виртуальных сценах, создавая эффекты динамических теней и отражений;

4. Предоставление программного интерфейса. Remote API [7] позволяет создавать пользовательские приложения и интегрировать их с Unreal Engine 5, что открывает широкие возможности для разработки собственных инструментов и приложений для виртуальной студии и виртуального видеопроизводства.

Таким образом, Unreal Engine 5 предоставляет все необходимые инструменты и возможности для создания высококачественной виртуальной студии и производства видеоконтента в реальном времени.

Для программной реализации приложения был выбран язык Python ввиду его простоты и удобства для трекинга PTZ-камер и отправки HTTP запросов в Unreal Engine 5. Для трекинга реальных камер была применена асинхронная библиотека Python с поддержкой протокола ONVIF [8]. Управление виртуальными камерами в сценах UE5 осуществляется через встроенный в движок Remote API.

Алгоритм создания решения для виртуальной студии выстроен следующим образом:

1. Создание виртуальной сцены внутри Unreal Engine 5

- 1.1 Настройка масштаба объектов в виртуальном мире

- 1.2. Настройка параметров освещения в сцене

2. Установка и настройка виртуальных камер в сцене, указание путей к виртуальным камерам в программе

3. Запуск встроенного веб-сервера UE5

4. Настройка плагинов Off World Live [9], Remote API Control в проекте

- 4.1. Установка виртуального устройства передачи потока камеры UE5

- 4.2. Запуск трансляции с виртуальных камер через плагин Off World Live

5. Установка PTZ-камер в правильное положение в студии, начало трекинга

- 5.1. Добавление в программу адресов PTZ-камер

6. Установка плагина NDI [10] source в OBS

- 6.1. Добавление PTZ и виртуальной камеры в качестве источников в OBS

- 6.2. Настройка фильтрации зеленого фона в потоке с виртуальной камеры

7. Склейка потоков в OBS. Вывод финальной картинке (рис.1).



Рис.1. Демонстрация работы

Таким образом, получается приложение, обладающее всеми необходимыми функциональными возможностями для создания и управления виртуальной студией и производства видеоконтента в реальном времени. Используя игровую среду Unreal Engine 5 в сочетании с языком программирования Python, данное приложение обеспечивает реалистичные 3D сцены, поддерживает работу с виртуальными и реальными камерами, а также предоставляет широкие возможности настройки освещения и интеграции пользовательских приложений через программные сетевые интерфейсы.

Такое приложение представляет собой инструмент, который сочетает в себе простоту управления, высокую производительность и минимальные финансовые затраты, что делает его идеальным выбором для создания виртуальных студий и производства видеоконтента для широкого спектра целей, включая образовательные, развлекательные и коммерческие проекты.

Заключение

В ходе работы были проанализированы различные существующие технологии реализации виртуального видеопроизводства и выявлены их преимущества и недостатки относительно критериев, указанных в разделе “Предлагаемая разработка”. Выводом явилось то, что среди программных реализаций не оказалось решения, способного обеспечить высокое качество съемки с минимальной задержкой по времени трекинга камеры. Физические аналоги оказались дорогостоящими и не поддерживающими трекинг более одной камеры с помощью одного устройства. Этократно увеличивает затраты на видеопроизводство при условиях студийной съемки с использованием нескольких камер. Разработанное решение позволяет минимизировать затраты на про-

изводство, при этом предлагая независимость работы от внешних ресурсов и возможность одновременного трекинга сразу нескольких PTZ-камер в пределах допустимого критериями разработки времени задержки в 100 - 200 мс. Резюмируя все перечисленные факторы, можно отметить, что разработка имеет хорошие шансы на практическое применение и способна достойно себя зарекомендовать на рынке, при этом имея большой потенциал развития и улучшения технологии.

Список литературы

1. An D. Technology-driven Virtual Production //Revista FAMECOS. – 2022. – Т. 29. – №. 1. – С. e43370-e43370.
2. Unreal Engine 5 URL: <https://www.unrealengine.com/en-US/unreal-engine-5> (дата обращения: 09.02.2024)
3. Шишалов И. С. и др. Способ и система определения параметров калибровки PTZ камеры. – 2019.
4. Chroma Key Filter URL: <https://obsproject.com/kb/chroma-key-filter> (дата обращения: 09.02.2024)
5. Antilatency Virtual Production URL: <https://virtualproduction.antilatency.com/> (дата обращения: 11.02.2024)
6. VIVE Tracker URL: <https://www.vive.com/cis/accessory/vive-tracker/> (дата обращения: 11.02.2024)
7. Remote Control API HTTP Reference URL: <https://docs.unrealengine.com/4.27/en-US/ProductionPipelines/ScriptingAndAutomation/WebControl/Endpoints/> (дата обращения: 16.02.2024)
8. ONVIF Core Specification URL: <https://www.onvif.org/specs/core/ONVIF-Core-Specification.pdf> (дата обращения: 16.02.2024)
9. OWL Live-Streaming Toolkit URL: <https://offworld.live/products/unreal-engine-live-streaming-toolkit> (дата обращения: 16.02.2024)
10. Using NDI with OBS URL: <https://streamgeeks.us/using-ndi-with-obs/> (дата обращения: 18.02.2024)

РАЗРАБОТКА СЕРВИСА ПРОКСИРОВАНИЯ ONVIF PTZ

Сенников М.В., Елисеенко А.М.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Работа посвящена разработке сервиса проксирования сетевого трафика по протоколу ONVIF (Open Network Video Interface Forum) с возможностью переключения между управляемыми устройствами, объединенными в группы. Целью работы является создание прокси-сервера, способного поддерживать подключение с программными и аппаратными

ONVIF-клиентами и перенаправлять запросы на нужную группу управляемых устройств, подменяя данные авторизации, URL видеопотоков и прочие данные, необходимые для стабильной работы ONVIF-клиентов. Также существует возможность управления RTSP-потоками, транслируемыми ONVIF-совместимыми устройствами.

Введение

В силу специфики работы распределенной сетевой телестудии, которой является Медиацентр МИЭМ, в течение одного рабочего дня съемки могут происходить в нескольких залах одновременно, а перерывы на перенастройку оборудования могут быть недостаточно длинными по времени, поэтому необходимость подключения пульта к новой группе камер создаёт дополнительные проблемы.

В ходе данной работы решается проблема переключения путем создания прокси-сервера, осуществляющего трансляцию трафика с устройств управления на группы камер с возможностью переключения между этими группами. Оператор имеет возможность предварительно задать группы камер в базе данных, ассоциируя "камера - порт прокси", а затем привязать пульт управления к прокси-серверу. Для переключения между группами достаточно воспользоваться API сервиса.

В процессе съемок несколько RTSP-клиентов могут получать изображение с камеры, при этом каждый поток кодируется отдельно. В связи с недостатком ресурсов камера может начать снижать качество потоков. Кроме того, все потоки передаются по одному каналу, а камер в сети может быть несколько, поэтому, переложив задачу дублирования потоков на устройство, расположенное в подсети потребителя, можно разгрузить основную и повысить стабильность всей системы. Поэтому дополнительно в сервис будет включен хаб для RTSP-потоков, что позволит снизить нагрузку на вычислительные ресурсы камер и сеть.

Обзор протоколов управления видеокameraми

Современное видеопроизводство и видеонаблюдение используют разные протоколы для управления устройствами, такие как:

- ONVIF [1]: стандарт открытого интерфейса сетевого видеонаблюдения, который обеспечивает совместимость между устройствами видеонаблюдения различных производителей. Протокол ONVIF позволяет управлять видеокameraми, получать видеопотоки, проводить аналитику и другие функции.
- VISCA [2]: протокол управления для профессиональных видеокameraх таких производителей, как Sony. Он позволяет управлять различными аспектами видеосъемки, такими как фокусировка объектива, экспозиция, цветопередача и другие параметры.
- Pelco-D/P [3, 4]: протоколы управления видеокameraми, разработанные компанией Pelco (теперь входит в Schneider Electric). Они обеспечивают возможность управления различными функциями камер, такими как панорамирование, наклон и масштабирование.

Среди вышеописанных протоколов ONVIF поддерживается наибольшим количеством устройств, а также является свободным, поэтому именно он был выбран для проекта.

Анализ существующих решений

Существует несколько решений, позволяющих проксировать ONVIF PTZ (Pan-Tilt-Zoom), наиболее распространенные представлены в таблице. Решения сравнивались по следующим метрикам:

- Способность проксирования ONVIF и RTSP – основной функционал разрабатываемого сервиса;
- Способность дублировать RTSP-потоки для освобождения вычислительных ресурсов камеры;
- Горизонтальная расширяемость. Подразумевает возможность запускать несколько экземпляров на разных устройствах (что актуально для проксирования RTSP), но обеспечивать единую панель управления [];
- Поддержка работы с несколькими соединениями на один экземпляр сервиса;
- Возможность переключаться между камерами “на лету”.

Таблица 1. Сравнение аналогов

| Решение | Проксирование ONVIF | Проксирование RTSP | Хаб RTSP |
|-------------------------------|---------------------|--------------------|----------|
| Happytimesoft onvif-proxy [6] | Да | Да | Да |
| BreeZe/rpos [7] | Да | Да | Да |
| Nginx [8] | TCP-прокси | Плагин | Плагин |

Таблица 2. Сравнение аналогов

| Решение | Горизонтальная расширяемость | Поддержка нескольких соединений | Поддержка комнат |
|-------------------------------|------------------------------|---------------------------------|------------------|
| Happytimesoft onvif-proxy [6] | Нет | Да | Нет |
| BreeZe/rpos [7] | Нет | Нет | Нет |
| Nginx [8] | Нет | Да | Нет |

Особенности PTZ при проксировании

Протокол ONVIF использует в качестве транспорта SOAP поверх HTTP [1]. Запрос представляет из себя XML, включающий данные авторизации и необходимые данные, ответ содержит подобный XML [9], который может иметь специфичные для конкретной камеры данные, такие как ссылки на сер-

висы и потоки, профили и прочие. Поскольку одной из задач является обеспечение максимально бесшовного переключения, необходимо подменять эти данные “на лету”, подставляя заранее сконфигурированные. При этом оператор получает возможность переключаться между камерами, имеющими разные пароли и адреса, без изменения настроек ONVIF-клиента.

Архитектура приложения

При разработке проекта был сделан упор на горизонтальную расширяемость приложения, так как проксирование RTSP-потоков в некоторых случаях требует больших вычислительных ресурсов. Для каждого соединения запускается отдельная инстанция slave-сервиса, содержащая ONVIF-прокси, написанный на python, и использующая MediaMTX [10] в качестве RTSP-прокси.

API реализуется в master-сервисе, который конфигурирует все инстанции по запросу. Поддерживается как REST, так и управление по WebSocket; в последнем случае все подписанные клиенты также получают нотификации об изменениях состояния системы в режиме реального времени.

Для хранения конфигурации по-умолчанию используется PostgreSQL, но может быть применена и другая СУБД, поддерживаемая SQLAlchemy [11]. На старте slave-сервис не имеет информации о том, какой именно сущности он соответствует; эту информацию он получает от master-сервиса, когда последний обращается к нему по адресу, записанному в БД.

Дистрибуция происходит в виде docker-compose проекта. Для запуска необходимо указать в нем необходимое количество slave-инстанций для данного устройства, а также порты. При этом важно, чтобы они совпадали с указанными в БД, так как не могут быть сконфигурированы автоматически.

Общая схема сервиса представлена на рисунке.

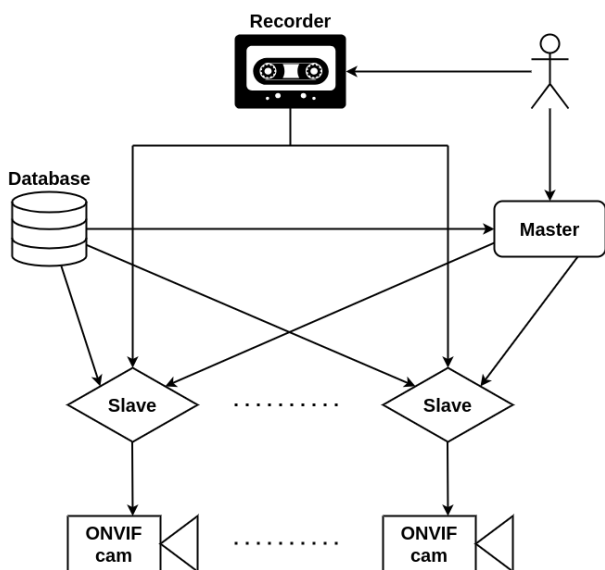


Рис. 1. Общая схема сервиса

Технические характеристики

Решение тестировалось на сервере с AMD Ryzen 5 5500 и 16 Гб ОЗУ (наличие графического процессора не играет роли). В таблицах ниже представлены результаты RPS тестов для двух запросов: GetProfiles, как одного из наиболее тяжелых, и GetSnapshotUri, как одного из наиболее легких.

Таблица 3. RPS тест GetProfiles

| | Камера | Прокси | Камера | Прокси |
|-----------------------------|--------|---------------|--------|---------------|
| Число запросов | 1000 | 1000 | 1000 | 1000 |
| Число конкурентных запросов | 1 | 1 | 20 | 20 |
| Время (с) | 121.4 | 124.0 (+2.1%) | 107.3 | 107.7 (+0.4%) |
| RPS | 8.24 | 8.07 | 9.32 | 9.29 |

Таблица 4. RPS тест GetSnapshotUri

| | Камера | Прокси | Камера | Прокси |
|-----------------------------|--------|--------------|--------|--------------|
| Число запросов | 1000 | 1000 | 1000 | 1000 |
| Число конкурентных запросов | 1 | 1 | 20 | 20 |
| Время (с) | 96.9 | 87.4 (-9.8%) | 64.4 | 60.0 (-6.8%) |
| RPS | 10.32 | 11.44 | 15.53 | 16.67 |

Из таблиц видно, что использование прокси сервера дает незначительную дополнительную задержку, а узким местом системы является сама камера. Также Таблица 4 демонстрирует эффект кэширования.

Для управления приложением может использоваться как классический REST API, так и WebSocket API. В случае подключения по WebSocket, в дополнение к базовым функциям, клиент также получает ссылку нотификаций об изменении статусов соединений, доступности камер, или наличии ошибок в момент их регистрации, что является большим преимуществом в условиях профессиональных съемок. Для удобства пользователя также добавлен веб-интерфейс, главная страница которого представлена на рис.2.

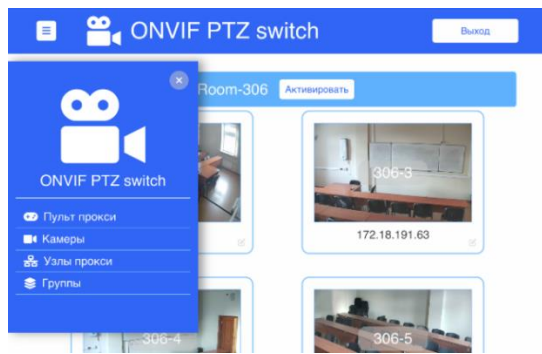


Рис.2. Главная страница сервиса

Заключение

Разработанный продукт позволяет эффективнее использовать дорогостоящее оборудование, а также время оператора, уменьшить количество человеческих ошибок, что важно в режиме повышенной психологической нагрузки во время съемок.

Список литературы

1. ONVIF Core Specification / Version 23.12; December, 2023 - с. 14-15.
2. VISCA Protocol [Электронный ресурс] Википедия. Свободная энциклопедия. – URL: https://en.wikipedia.org/wiki/VISCA_Protocol (дата обращения: 10.11.2023).
3. Pelco-D Protocol Specification / Version 2.12.0; Nov.09.2022.
4. Pelco-P Protocol Specification / Version 2.12.0; Nov.09.2022.
5. Костиков, Ю. А. Разработка информационной системы для распределенной обработки потоковых данных / Ю. А. Костиков, А. М. Романенков // Сборник научных трудов кафедры прикладной математики и программирования по итогам работы постоянно действующего семинара «Теория систем». – Москва: Федеральное государственное бюджетное образовательное учреждение высшего образования "Российский государственный университет имени А.Н. Косыгина (Технологии. Дизайн. Искусство)", 2022. – с. 57-59.
6. ONVIF Проху [Электронный ресурс] Happytimesoft – URL: <https://happytimesoft.com/products/onvif-proxy/index.html> (дата обращения: 17.11.2023).
7. BreeZe/rpos [Электронный ресурс] GitHub – URL: <https://github.com/BreeZe/rpos> (дата обращения: 17.11.2023).
8. Nginx [Электронный ресурс] URL: <https://nginx.org> (дата обращения: 17.11.2023).
9. Simple Object Access Protocol (SOAP) 1.1 [Электронный ресурс] URL: <https://www.w3.org/TR/2000/NOTE-SOAP-20000508> (дата обращения: 19.11.2023).
10. MediaMTX [Электронный ресурс] GitHub – URL: <https://github.com/bluenvion/mediamtx> (дата обращения: 12.01.2024).
11. SQLAlchemy 2.0 Documentation // Engine Configuration [Электронный ресурс] URL: <https://docs.sqlalchemy.org/en/20/core/engines.html> (дата обращения: 10.12.2023).

КОРРЕКТИРУЮЩИЕ ПОДХОДЫ К ВИЗУАЛИЗАЦИИ ГЕОЛОЦИРОВАННЫХ ОБЪЕКТОВ МАРШРУТИЗАЦИИ НА КАРТОГРАФИЧЕСКИХ ПОДЛОЖКАХ

Егоров А.М.¹, Подгук А.Е.²,
Половцев А.В.¹, Чертов В.Д.²

Национальный исследовательский университет

«Высшая школа экономики»,

¹ департамент прикладной математики,

² департамент компьютерной инженерии

МИЭМ НИУ ВШЭ

Аннотация

Сегодня Интернет – это не только поле конкурентной борьбы операторов сети за новых пользователей посредством улучшения качества предоставляемого сервиса, но и стратегически важная задача для организаций, регулирующих деятельность в Интернете и операторов телекоммуникационных сетей, озабоченных проблемами быстрой, надёжной и безопасной доставки потребителю информации, циркулирующей в том или ином региональном сегменте сети [1]. Для мониторинга текущей ситуации весьма желателен набор инструментов, облегчающий анализ и оптимизацию происходящих в сети процессов. В качестве одного из таких инструментов рассматривается средство визуализации данных об объектах маршрутизации. В данной работе рассматриваются два основных подхода к отображению объектов маршрутизации на географической карте, а также описывается процесс формирования уточнённой базы данных, содержащей информацию о геокоординатах объектов маршрутизации.

Введение

Практически у любого оператора сети/провайдера телекоммуникационных услуг одной из наиболее актуальных является задача сокращения времени доставки контента до пользователя. Для решения данной проблемы была разработана технология CDN. CDN (Content Delivery Network) – это географически распределённая сетевая инфраструктура, обеспечивающая быструю доставку контента пользователям веб-сервисов и сайтов. «Логистическим центром» CDN является система маршрутизации запросов. В задачу этой системы входит выбор оптимального пути до граничного узла для максимально быстрого обслуживания запроса клиента. Эффективность сети CDN во многом определяется тем, насколько близко к потребителю контента располагаются точки ее присутствия (узлы) [2]. Для определения степени близости узла к пользователю необходимо понимать, как относительно друг друга расположены объекты маршрутизации и сам пользователь, то есть нужно знать их геокоординаты. Объектом маршрутизации в контексте данной работы именуется пара «номер автономной системы/префикс». Номер автономной системы определяет множество сетей с единой политикой маршрутизации, а префикс определяет некоторое непрерывное конечное множество сетевых адресов. Таким образом, основной про-

блемой визуализации геолоцированных объектов маршрутизации является отображение множества объектов и связей между этими множествами.

Кроме того, решение этой задачи требует формирования корректной базы данных геопривязки, основываясь на максимально достоверных для конкретной территории исходных данных. В мировой практике уже существует подобное решение - база данных GeoIP от компании MaxMind (США). Но основной проблемой использования GeoIP и подобных существующих баз данных является весьма существенная погрешность при определении координат и ошибки геопривязки. Поэтому на некоторых территориях, в частности расположенных на евразийском континенте, существующие базы данных имеют существенные проблемы с точностью геопривязки IP-адресов, отрицательно сказывающиеся на формировании маршрутных политик и в конечном счёте негативно влияющие на скорость доставки контента.

Однако, главной трудностью остаётся реализация корректного отображения множества объектов маршрутизации на географическую карту сетевой инфраструктуры.

За последние 15 лет было осуществлено множество попыток решения этой проблемы. Наиболее значимыми из них были RIPE Geoloc (2011 год), RIPE IPmap (2018-2019), введение атрибута «Geofeed» в базу данных регионального реестра Азиатско-Тихоокеанского региона APNIC и ряд других. Инженеры и исследователи отталкивались в своих решениях от природы сетевого взаимодействия и топографии и попадали «в ловушку» абстрактной визуализации, которая не позволяла решать практические задачи управления сетью, управления клиентами, управления взаимодействием с регуляторами.

В представленной исследовательской работе в основу положены конкретные задачи визуализации объектов маршрутизации, в ходе решения которых был осуществлён переход от классической топографии к карте административно-территориального деления.

Это является новым подходом к отображению геолоцированных объектов маршрутизации на картографических подложках.

В процессе проведённого исследования сформирована уточнённая база данных геопривязки объектов маршрутизации и реализованы два способа их визуализации на географической карте РФ.

Формирование уточнённой базы данных геопривязки

На первом этапе исследования сформирована база данных с поддержкой механизма отслеживания истории изменения атрибутов записей, используемых для хранения уточнённых данных об объектах маршрутизации на территории РФ, набор скриптов, выполняющих добавление новых записей в базу, а также пользовательский интерфейс формирования запросов к БД с целью получения данных из нее для отображения на картографической подложке.

В качестве основного источника данных используется интернет-регистратура RIPE NCC, которая распределяет интернет-ресурсы на территории Европы, Ближнего Востока и Центральной Азии.

Для проверки правильности отрисовки точек (автономных систем) и линий (связей между автономными системами) на карте РФ применена предоставленная организацией-партнёром тестовая база данных, содержащая информацию о 100 автономных системах, в т. ч. их географических координатах, а также таблица связей между автономными системами.

Основные данные об объектах маршрутизации содержатся в базе данных RIPE DB [3], где требуется для проведения исследования информацией были данные об объектах route и автономных системах, к которым привязаны route (рис.1). Объекты route одни из самых важных объектов в RIPE DB, они связывают автономную систему (AS) с блоками IP-адресов и их префиксами. Префикс определяет количество IP-адресов в разделе конкретного узла сети.

| | |
|----------------|----------------------|
| route: | 194.113.52.0/23 |
| descr: | Beumer |
| origin: | AS30941 |
| mnt-by: | DTS-MNT |
| created: | 2004-02-23T09:43:17Z |
| last-modified: | 2004-02-23T09:43:17Z |
| source: | RIPE# Filtered |

Рис.1. Пример объекта Route в RIPE DB

Адреса данных блоков IP получены с помощью RIPE ATLAS API и модуля whois [4].

RIPE ATLAS использует сведения, поступающие с датчиков, закрепленных за определенными автономными системами. А в них соответственно содержится информация о стране и координатах нахождения данного IP-блока (рис.2).

```
HTTP 200 OK

Allow: GET, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

{
  "count": 1,
  "next": null,
  "previous": null,
  "results": [
    {
      "address_v4": "62.150.88.222",
      "address_v6": null,
      "asn_v4": 9155,
      "asn_v6": null,
      "country_code": "RU",
      "description": "Qualitynet General Trading and Contracting Company (Prope)",
      "firmware_version": 4780,
      "first_connected": 1525761517,
      "geometry": {
        "type": "Point",
        "coordinates": [
          47.9785,
          29.3805
        ]
      }
    }
  ]
}
```

Рис.2. Пример данных, полученных через RIPE ATLAS API

Однако не все IP-префиксы, объявленные в RIPE DB, можно обнаружить через датчики в RIPE ATLAS [5]. Для устранения подобных пробелов в данных, а также для сравнения с уже имеющимися адресами, использовался модуль whois, с помощью которого можно получить адрес, где зарегистрирован данный объект.

Таким образом, через координаты или адрес определяется регион, где расположен данный блок, который в свою очередь отображается на карте.

Визуализация объектов маршрутизации

Визуализация объектов маршрутизации на картографической подложке с помощью базы данных геопривязки позволяет наглядно представить сложные сетевые структуры и процессы передачи данных, облегчая понимание и анализ маршрутной информации, например, в контексте выявления нарушений в работе автономной системы (АС) [6]. Точность геолокации может заметно варьироваться в зависимости от используемой базы данных. Мобильные устройства создают дополнительные сложности, поскольку их IP-адреса могут изменяться в зависимости от текущего местоположения пользователя. Все эти факторы влияют на географическую точность геолоцирования при визуализации автономных систем и других объектов маршрутизации на карте. В процессе проведенного исследования реализованы два подхода к отображению объектов маршрутизации на географической карте.

Отображение объектов маршрутизации в виде точек на карте

Рассмотрим отображение автономных систем в виде точек на географической карте РФ. Такой подход помогает понять, к какой автономной системе могут относиться IP адреса, находящиеся в рассматриваемом регионе. Также в этом случае удобно отображать статистику по конкретным маршрутизаторам, что может помочь оператору, например, в поиске конкретного неисправного узла сети. В большинстве случаев одному оператору принадлежат несколько автономных систем, поэтому при отображении на карте они образуют кластеры. Такой подход к визуализации может быть приемлемым, если отображаемых объектов немного (в этом случае рассматриваются исключительно автономные системы), иначе на карте будет избыточное количество точек, из-за чего зрительное восприятие информации затрудняется. Кроме того, при таком подходе визуализация не даёт полного представления о рассматриваемых узлах сети, поскольку, например, к IP-префиксам и автономным системам не применимо понятие точного местоположения, а сами IP-адреса часто могут менять своё истинное местоположение.

В ходе исследования были реализованы функции отрисовки точек и линий на карте, кластеризации точек и вывода на экран информации о каждой автономной системе (рис.3).

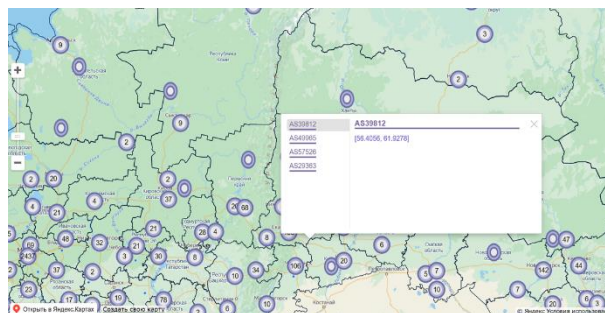


Рис.3. Отображение автономных систем в виде точек на карте

Отображение объектов маршрутизации по принадлежности к регионам

При данном формате отображения объектов маршрутизации аналитики сетевых данных могут оперативно отслеживать инциденты и решать проблемы, касающиеся различных маршрутизаторов, которые могут возникнуть в конкретном регионе. Этот подход позволяет собирать статистику не по конкретным узлам сети, а по всем узлам сети заданного региона, что предоставляет возможность оценивать гораздо больший объём информации о состоянии и уровне защищенности объектов маршрутизации конкретного региона. Также в этом случае можно сравнительно легко узнать в каком регионе располагается заданный IP/префикс/автономная система. Это может быть полезно для получения информации о том, как организован трафик до конечного пользователя, а также для определения региона, в котором зарегистрирован тот или иной IP-адрес. Очевидно, что при рассмотренном подходе отображение выглядит менее нагруженным и визуально более понятным для пользователя, но, в сравнении с предыдущим случаем, мы некоторым образом жертвуем географической точностью (рис.4).

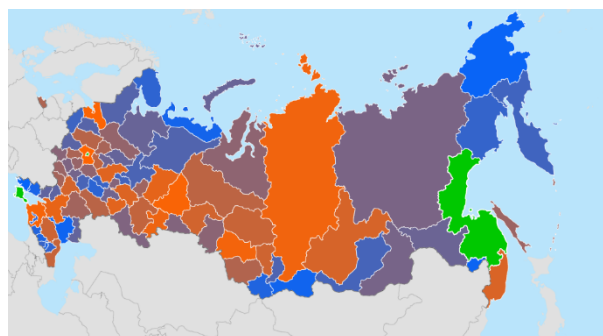


Рис.4. Отображение объектов маршрутизации по принадлежности к регионам

Заключение

Результаты проведенного исследования позволяют предложить каждый из рассмотренных способов отображения геолоцированных объектов маршрутизации на картографических подложках как средство поддержки управленческих решений для сетевых операторов. При этом для их совместного использования может быть разработана многослойная струк-

тура визуализации, поддерживающая отслеживание изменений базы данных во времени и позволяющая оценивать ситуацию в сети с различных позиций.

Список литературы

1. Робачевский А. Интернет изнутри: Экосистема глобальной Сети. – М.: Альпина Паблицер, 2017. – 272 с.
2. CDN: что такое и как это работает // URL: <https://selectel.ru/blog/review-cdn/>
3. Основы работы с базой данных RIPE // URL: <https://habr.com/ru/companies/Linx/articles/526508/>
4. Ipvhois // URL: <https://pypi.org/project/ipwhois/>
5. Geolocation in the RIPE Database // URL: <https://apps.db.ripe.net/docs/Related-Software-and-Tools/Geolocation-in-the-RIPE-Database/> (дата обращения: 01.02.2024).
6. Hal Roberts, David Larochelle Mapping Local Internet Control // Berkman Center for Internet & Society Harvard University - URL: <https://cyber.harvard.edu/netmaps/mlc.pdf> (дата обращения: 07.02.2024).

ПРИМЕНЕНИЕ ТЕОРИИ ПРИНЯТИЯ РЕШЕНИЙ В ЗАДАЧЕ ВЫБОРА СТРАТЕГИИ РЕАГИРОВАНИЯ НА ИТ-ИНЦИДЕНТЫ

Омельченко К.Г., Стрельников Д.Ю.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В работе рассматривается текущее состояние проблемы управления инцидентами в информационных системах. Также изучается возможность применения методов теории принятия решений для автоматизированного выбора наиболее рациональной стратегии реагирования на инциденты в информационных системах.

Введение

С ростом сложности современных информационных систем все более актуальным становится вопрос сокращения времени, требуемого для решения инцидентов, то есть незапланированных событий, которые могут привести к снижению качества работы системы [1].

Данная работа нацелена на изучение возможности повышения эффективности процесса управления инцидентами и на его автоматизацию за счет применения теории принятия решений в задаче экспертного выбора стратегии по устранению инцидента.

Существующие методологии и решения

На сегодняшний день можно выделить две наиболее популярные методологии, описывающие подход к управлению процессами, связанными с информационными технологиями (ИТ), и включа-

ющими в себя работу с инцидентами – ITIL [2] и DevOps [3].

Библиотека инфраструктуры информационных технологий (ITIL) содержит набор рекомендаций и правил по эффективному согласованию ИТ-процессов. Одним из ключевых компонентов ITIL является управление инцидентами, включающее в себя несколько этапов от идентификации появления инцидента до его решения. Однако, согласно статистике [4], требуется улучшение данного компонента с точки зрения эффективности бизнес-процессов, касающихся решения инцидентов.

В отличие от ITIL, где инциденты и вызвавшие их проблемы рассматриваются отдельно, методология DevOps подразумевает необходимость их совместного рассмотрения [5]. Также, методология указывает на возможность существования сразу нескольких причин возникновения инцидентов. Управление инцидентами в DevOps подразумевает, что решать инцидент должен наиболее осведомленный о работе системы или вышедшего из строя компонента сотрудник. Такой подход может быть эффективен в случае небольших систем, но в масштабных проектах знания о системе обычно распределены между разными командами, что может усложнить коммуникацию между экспертами в процессе выбора стратегии.

В рамках обеих методологий существуют специальные инструменты, облегчающие и систематизирующие некоторые этапы управления инцидентами. Так, Service Desk [6] позволяет подробно описать инциденты и принимает запросы на их обслуживание, CMDB [7] выступает в роли хранилища, включающего все элементы ИТ-инфраструктуры организации и логические связи между ними. В качестве средства автоматизации процесса устранения инцидентов могут выступать IRP-системы [8], позволяющие решать типовые инциденты по заранее заданным сценариям, основанным на имеющихся данных по уже решенным инцидентам. Стоит заметить, что такие сценарии не всегда эффективны, так как в комплексных ИТ-системах причины возникновения инцидента могут быть разными при похожих внешних проявлениях.

Проблемы, которые потенциально могли привести к рассматриваемому инциденту, иногда достаточно трудно идентифицировать с высокой точностью и в полном объеме, их можно только предположить. Это вносит элемент прогнозирования для решения типовых и уникальных инцидентов в рамках оптимизации этого процесса. Поэтому представляется возможным применение методов поддержки принятия решений для выбора стратегии реагирования на инциденты. Следует отметить, что такие методы могут использоваться и на других этапах управления инцидентами, например, на этапе приоритизации [9].

Анализ применимости методов

Выбор стратегий реагирования на инциденты должен производиться с учетом ряда критериев, зачастую имеющих взаимопротиворечивый характер

[10]. Соответственно, целесообразно рассматривать использование методов теории принятия решений. Также инциденты могут решаться не только одним экспертом (техническим специалистом), но и группой экспертов, что тоже накладывает ограничения на использование определенных методов.

Рассмотрим наиболее подходящие для решения такой задачи методы.

Метод ранга основывается на шкале оценивания, которая используется несколькими экспертами для присвоения оценок определенным альтернативам. Сумма оценок по альтернативам с последующим нормированием относительно общей суммы позволяет определить ранг варианта. Метод требует проверки согласованности на основе расчета дисперсии оценок каждого эксперта и каждой альтернативы [11].

Метод предпочтений основан на ранжировании альтернатив каждым из экспертов по предпочтению в порядке убывания важности альтернативы. Полученные оценки заносятся в матрицу, после чего происходит ее преобразование и рассчитываются веса каждой альтернативы относительно общей суммы оценок. В итоге, чем больше вес, тем более важна альтернатива. Для исключения резкого различия во мнениях экспертов метод позволяет проверить согласованность на основе коэффициента конкордации [11].

И метод ранга, и метод предпочтений подходят для ранжирования стратегий устранения инцидента в группе экспертов. Применение данных методов может помочь систематизировать процесс коммуникации между экспертами и упростить групповую оценку стратегий.

Метод попарного сравнения – индивидуальный метод, основанный на сравнении каждой пары альтернатив на основе шкалы оценивания. По полученным оценкам строится матрица альтернатив и подсчитываются веса каждой из них на основе средних геометрических значений строк матрицы. Соответственно, больший вес означает большую ценность альтернативы. Недостатком метода является рост сложности оценивания пар экспертом с ростом числа альтернатив [12]. Во избежание противоречий в оценках проводится проверка согласованности [11]. Метод может быть полезен для выбора стратегий, поскольку их описание может быть достаточно объемным, что вносит определенные трудности в вопрос оценки и ранжирования стратегий.

Метод перестановок использует процедуры назначения весов критериев и попарного сравнения оценок предпочтительности альтернатив по каждому критерию, на основе чего находится лучшая альтернатива. Недостатком метода является резкий рост количества перестановок при увеличении числа альтернатив. Также метод обладает высокой чувствительностью к случайным ошибкам эксперта, а сама процедура задания весов может быть сложной для понимания экспертом, не имеющим специальной подготовки в области теории принятия решений [12].

Метод аналитических иерархий позволяет оценить альтернативы на основе множества критериев.

Метод основан на определении весов критериев, а затем весов альтернатив за счет их попарного сравнения. Ценность альтернативы определяется на основе суммы произведений весов для каждой альтернативы. Преимуществом метода является возможность применения в случае, когда эксперты не могут дать абсолютных оценок альтернатив по критериям [12], что характерно для задачи решения инцидентов. Однако, это может означать некоторую оторванность от предпочтений эксперта и требует наличия транзитивности для качественных величин [13]. Расчет согласованности позволяет выявить ошибки эксперта при заполнении матрицы попарных сравнений.

Метод платежной матрицы позволяет гибко определить наиболее подходящую альтернативу на основе вычисления различных характерных оценок, полагаясь на величину затратности каждой стратегии, исходя из какого-либо предшествующего события или исходной информации. Характерные оценки могут выражать максимальные и минимальные затраты, среднеарифметические или перерасход [12]. Так, в дальнейшем, эти оценки применяются для вычисления критериев выбора решений, как, например, Севиджа, Вальда или Лапласа, которые указывают на вариант с минимаксным риском, минимаксными затратами и минимумом среднеарифметических затрат, соответственно. Платежная матрица позволяет в ходе решения инцидента оценить стратегии с различных сторон с точки зрения рисков их выполнения, что важно в условиях неопределенности и высокой ответственности.

Методы перестановок, аналитических иерархий и платежной матрицы могут быть положены в основу процедур оценивания и выбора стратегий реагирования при совместном рассмотрении инцидента и вызвавших его проблем, что подразумевается методологией DevOps, при невозможности однозначно определить все причины возникновения инцидента.

Как было указано ранее, методы теории принятия решений предполагают способ вычисления коэффициента согласованности. Возможность получения согласованных оценок – один из важнейших показателей эффективности метода, особенно в ходе коллективного решения [11], поэтому он особенно важен в контексте решения инцидентов, чтобы минимизировать нежелательные траты компании, предоставляющей ИТ-услугу или владеющей ИТ-системой. Если коэффициент оказывается ниже заданного порога, как, например, в случае метода предпочтений это 0.5, то требуется повторить выполнение метода.

Заключение

Таким образом, рассмотренные методы теории принятия решений наиболее целесообразны для использования в процессе управления инцидентами не только на этапе приоритизации, но и на этапе выбора стратегии устранения инцидента. Причем на их основе возможно построить комплексную автоматизированную систему поддержки принятия решений с взаимной интеграцией методов, чтобы не только нивелировать недостатки каких-либо из них, но и вос-

пользоваться их преимуществами для решения задачи выбора стратегий реагирования при различном количестве экспертов, возможных стратегий, причин и проблем, вызвавших инцидент.

Список литературы

1. Сиднев С. М., Острейковский В. А. Обзор и поиск решений оптимизации времени разрешения ИТ-инцидентов // НАУКА И ИННОВАЦИИ XXI ВЕКА. – 2016. – С. 64-69.
2. AXELOS Limited and Stationery Office. ITIL Foundation, ITIL 4 Edition. – TSO (The Stationery Office), 2019.
3. Bass L., Weber I., Zhu L. DevOps: A software architect's perspective. – Addison-Wesley Professional, 2015.
4. Danby, S. ITSM Statistics: The 2023 State of ITSM Process Adoption. – URL: <https://blog.invgate.com/itsm-statistics> (дата обращения: 29.01.2024).
5. Atlassian. Управление проблемами и управление инцидентами. – URL: <https://www.atlassian.com/ru/incident-management/devops/incident-vs-problem-management> (дата обращения: 03.02.2024).
6. Tang X., Todo Y. A study of service desk setup in implementing IT service management in enterprises. – 2013.
7. Sharifi M., Ayat M., Sahibudin S. Implementing ITIL-based CMDB in the organizations to minimize or remove service quality gaps // 2008 Second Asia International Conference on Modelling & Simulation (AMS). – IEEE, 2008. – P. 734-737.
8. Очередыко А. Р. и др. Исследование ирп-систем на основе анализа механизмов реагирования на инциденты информационной безопасности // Прикаспийский журнал: управление и высокие технологии. – 2021. – №. 1 (53). – С. 74-82.
9. Jantscher M., Schwarz C., Zinser E. Decision support in it service management: Applying ANP methodology to the itil incident management process // International Journal of the Analytic Hierarchy Process. – 2015. – V. 7. – №. 1.
10. IT Expert. Описание ключевых процессов управления ИТ-услугами. – URL: <https://www.itexpert.ru/rus/TTEMS/proces/> (дата обращения: 04.02.2024).
11. Смородинский С. С., Батин Н. В. Системный анализ и исследование операций: лабораторный практикум для студентов специальности “Автоматизированные системы обработки информации”. – 2009.
12. Вишнеков А. В. Методы принятия проектных решений в CAD/CAM/CAE системах электронной техники // М.: МГИЭМ. – 1999.
13. Коробов В. Б., Тутыгин А. Г. Преимущества и недостатки метода анализа иерархий // Известия Российского государственного педагогического университета им. АИ Герцена. – 2010. – №. 122.

ПРЕДЛОЖЕНИЕ СТРУКТУРЫ МОДЕЛИ СИСТЕМЫ СВЯЗИ ДЛЯ ВЫСОКОСКОРОСТНЫХ ЖЕЛЕЗНОДОРОЖНЫХ МАГИСТРАЛЕЙ (ВСЖМ)

*Никитин С.Е., Клюев Н.С., Сахаутдинов Ш.Р.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе рассматривается структура модели беспроводной системы связи с ж/д составом на ВСЖМ, которая закладывает основу для создания математической модели для предсказания характеристик работы такой системы в различных сценариях. Данная работа выполнена в рамках выполнения 1-го этапа гранта НИУ ВШЭ «Исследование методов повышения помехозащищенности систем мобильной связи 4G/5G/6G, Wi-Fi, IoT с помощью сигнально-кодовых конструкций для сложных каналов в соответствии со Стратегией Развития Телекоммуникаций 2035».

Введение

С активным развитием отечественной железнодорожной отрасли и появлением высокоскоростных ж/д магистралей, высокая производительность в реальном времени становится основной проблемой в области обеспечения ж/д связи. Устаревшая технология GSM-R (2G) не подходит для обеспечения надежной связи «поезд-земля» и достижения требуемой скорости передачи данных на всем маршруте следования высокоскоростного состава, что требует развертывания и внедрения более актуальных технологий [1].

Структура модели, предлагаемая в данной работе, учитывает возможность построения системы беспроводной связи для ВСЖМ на основе технологий, потенциально доступных для обеспечения высокоскоростного доступа в Интернет для поездов:

1. Системы связи TSN (Track-Side-Network)
2. Мобильной системы связи LTE

При установлении соединения «поезд-земля» данные технологии могут использоваться независимо или совместно, обеспечивая высокую надежность соединения и высокую пропускную способность.

Сети LTE

В настоящий момент железнодорожная связь в основном обеспечивается устаревшей технологией GSM-R (2G), которая часто не позволяет достичь требуемой скорости передачи данных на состав. Решением может послужить постепенное развертывание мобильными операторами современных систем связи LTE.

LTE — это первый стандарт мобильной связи на базе IP-технологий с полной пакетной коммутацией, разработанный 3GPP [2]. В отличие от GSM-R, сеть LTE распределяет сетевые ресурсы между пользова-

телями и приложениями в зависимости от фактического спроса на передачу данных.

Так как в LTE передача данных происходит пакетами, была представлена упрощенная магистральная сеть под названием Evolved Packet Core (EPC) с меньшим количеством элементов, чем в предыдущих стандартах. Благодаря пакетированной передаче данных EPC может использовать любую транспортную сеть на базе IP [2], например, Carrier Ethernet.

На рис.1 приведена типовая структура сети LTE при её организации для обеспечения железнодорожной связи.

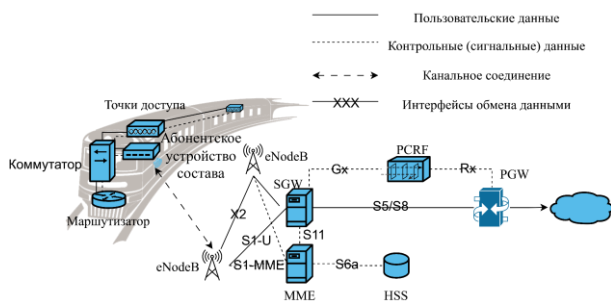


Рис.1. Типовая структура сети LTE при организации ж/д связи

Для LTE был разработан новый радиointерфейс, основанный на ортогональном частотном разделении множественного доступа (OFDMA) в нисходящем канале и множественном доступе с разделением каналов по частоте (FDMA) с одной несущей (SC-FDMA) в восходящем канале [2]. Схемы модуляции и кодирования в LTE выбираются динамически в зависимости от радиоусловий и спроса на трафик [3]. Все это обеспечивает гораздо более высокую спектральную эффективность, чем разработанные ранее стандарты мобильной связи [4]. Это обусловлено усовершенствованной модуляцией (OFDMA), мультиплексированием (до 64QAM), а также использованием множественного входа и множественного выхода для антенн базовых станций (MIMO).

Стоит отметить, что LTE может работать в различных полосах частот: 1,4 МГц, 3 МГц, 5 МГц, 10 МГц, 15 МГц или 20 МГц (и даже больше - с агрегацией несущих в LTE-Advanced). Такой диапазон частотных полос позволяет операторам гибко управлять доступным радиочастотным спектром. Например, оператор может разделить радиополосу, используемую сетью GSM (рефарминг спектра), где одна часть будет использоваться GSM, а в другой будет развернута новая сеть LTE. Выбирая более широкую или более узкую полосу LTE, оператор сети может решить, какую часть радиоресурсов выделить для LTE, а какую оставить для GSM [4].

Это может быть очень важной особенностью для железных дорог, поскольку это означает, что железнодорожная сеть LTE может быть развернута вместе с сетью GSM-R в том же диапазоне. Таким образом, переход на LTE будет постепенным. При организации связи «поезд-земля» базовые станции мобильных операторов подключаются к абонентскому

устройству, которые находятся на крыше поезда. В поезде абонентское устройство соединяется с маршрутизатором и точками доступа через коммутатор.

Сети TSN

TSN (Track-side-network) – это выделенные сети железнодорожных операторов. Такие сети включают в себя опорную централизованную сеть, подключенную к базовым станциям с использованием оптоволоконной, и бортовую сеть поезда. Абонентское устройство поезда связано с опорной сетью посредством радиоканала, который выстраивается между поездом и базовыми станциями, расположенными вдоль путей [7].

При проектировании сети TSN в зависимости от выбранной технологии, частотного диапазона и других условий, проводятся расчеты запаса энергии на линии между поездом и базовой станцией. Также определяется рекомендуемое расстояние между базовыми станциями. Далее происходит монтаж базовых станций вдоль путей и их подключение к централизованной опорной сети по оптоволокну.

При построении структуры модели сети TSN будем ориентироваться на уже существующее решение компании Максима Телеком [8] (рис.2).

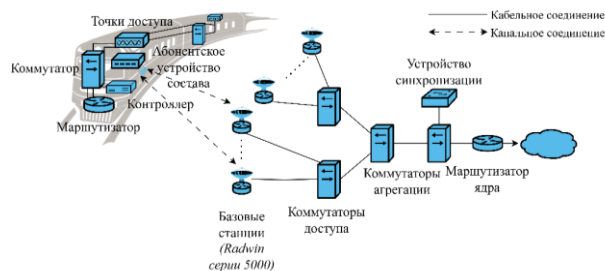


Рис.2. Структура сети TSN компании Максима Телеком

В головном вагоне состава расположено абонентское устройство для установления соединения с базовыми станциями, расположенными вдоль путей.

Для организации сети Wi-Fi в головном вагоне к коммутатору подключается абонентское устройство, маршрутизатор и первая точка доступа. Между вагонами проложены две кабельные трассы для повышения отказоустойчивости сети. Канальный уровень для пользовательского трафика терминируется на маршрутизаторе в головном вагоне.

Абонентское устройство состава посредством радиоканала соединено с базовыми станциями, соответствующими стандарту 802.11n. Передача данных между абонентским устройством и базовыми станциями происходит по проприетарному протоколу, основанному на TDM (Time Division Multiplexing).

Разрешенный частотный спектр разбивается на несколько непересекающихся каналов. К примеру, для Московского метрополитена разрешенный спектр 5150 – 5350 МГц разбивается на несколько подканалов по 40 МГц каждый. При этом базовые станции работают на различных каналах в последовательности 1-3-

5-2-4, чтобы избежать внесения помех в радиоканал каждой базовой станции соседними.

Базовые станции подключены к коммутаторам доступа посредством WDM (Waveless-division multiplexing) технологии, что позволяет использовать каждое волокно кабеля одновременно для приема и передачи данных.

Коммутаторы доступа подключаются в коммутаторы агрегации пропускной способности 1 Гбит/с с георезервированием на два uplink канала. В свою очередь, коммутаторы агрегации подключены в коммутаторы ядра, поддерживающие пропускную способность до 10 Гбит/с.

В коммутаторы ядра также подключаются устройства синхронизации, которые позволяют осуществить синхронизацию базовых станций по протоколу, схожему с RTP 1588v2, и маршрутизаторы, осуществляющие подключение к сети Интернет.

Деление по задачам

В сетях железнодорожной связи кроме данных, которые необходимы для контроля состояния состава передаётся и другая информация. Для корректного описания железнодорожных систем связи данные разбиваются на категории. К примеру, категорий может быть три [9]:

1. Системы контроля и отслеживания поезда (TCMS)

Включает функции управления и отслеживания местоположения поезда. Важным требованием к сетям, передающим информацию TCMS, является безопасная и надежная коммуникация между конечными устройствами.

2. Мультимедийные сервисы (OMTS)

Включают в себя все вспомогательные системы для нормальной эксплуатации поезда, кроме TCMS. Коммуникация в области OMTS не является непосредственно критичной для сохранности жизни пассажиров. К этой категории относится управление информационными дисплеями и системой информирования пассажиров.

3. Клиентоориентированные системы (COS)

Включают услуги, связанные с дополнительными удобствами пассажиров: индивидуальные мультимедийные дисплеи или услуги Wi-Fi. Такие системы не являются критически важными для безопасности.

Общая структура сети

Совместное двухчастотное использование сетей LTE и TSN с размещением базовых станций обеих технологий вдоль железнодорожных путей следования состава, с резервированием одной сетью другой, с суммированием трафика для пассажиров на подвижном составе позволит закрыть потребности в передаче данных на всех уровнях классификации.

Кроме того, совместное использование сетей позволяет эффективно управлять переходом между ними в зависимости от текущих условий связи. Например, при замираниях или прочих проблемах с установлением стабильного соединения в одной из сетей, автоматическое переключение на резервную позволяет сохранить непрерывность соединения и обеспе-

чивает надежный доступ во внешнюю сеть как для пассажиров, так и для систем контроля поезда.

Таким образом, общая структура сети связи должна учитывать установление соединения «поезд-земля» с подключением через оба вида сетей с резервированием ими друг друга. Её конечный вид, прежде всего, определяется выбранным способом построения сети TSN. Для структур сетей LTE и TSN, приведенных на рисунках 1 и 2, обмен технической информацией между сетями может происходить через промежуточный шлюз, подключенный к SGW для LTE и маршрутизатору ядра для TSN.

Создание математической (и, впоследствии, программной) модели такой сети – крайне науко- и ресурсоемкая задача, поскольку для ее реализации необходима разработка и программная реализация моделей каждой из сетей с учетом задержек и потери пакетов при передаче данных от базовых станций к ядру сети, многолучевого канала передачи, передачу данных в бортовой сети поезда. Тем не менее, создание модели является критически важным шагом к обеспечению стабильной железнодорожной связи на ВСЖМ.

Для её реализации необходимо, прежде всего, определить конечную структуру целевой системы связи, затем разработать имитационные математические модели сетей LTE и TSN в разных сценариях использования (как для канального, так и для системного уровней), и, наконец, создать программную модель в соответствии с разработанными математическими моделями сетей.

Моделирование должно быть произведено с учетом различных сценариев, которые отличаются скоростью состава, типом состава, погодными условиями, рельефа местности, электромагнитной обстановки и проч.

Заключение

Построение современной сети железнодорожной связи является критически важным условием как для безопасности пассажиров, так и для их комфорта во время поездки.

В работе были рассмотрены типовые структуры сетей LTE и TSN, описан общий принцип построения современной сети ж/д связи, предполагающей совместное использование сетей обоих типов.

Для определения конечной структуры сети необходимо определить способ построения сети TSN, поскольку реализация этих сетей не регламентируется международными стандартами. Тем не менее, в работе был предложен способ взаимодействия сетей LTE и TSN через промежуточный шлюз, который, впрочем, справедлив для конкретного способа построения сети, выбранного компанией Максима Телеком.

Список литературы

1. Y. Hu, H. Li, Z. Chang and Z. Han. Scheduling Strategy for Multimedia Heterogeneous High-Speed Train Networks // IEEE Transactions on Vehicular Technology, vol. 66, no. 4, pp. 3265-3279, April 2017, doi: 10.1109/TVT.2016.2587080.

2. M. Sauter. From GSM to LTE: an introduction to mobile networks and mobile broadband // John Wiley & Sons, first ed., 2011. ISBN 978-0470667118.

3. S. Sesia, I. Toufik, and M. Baker. LTE - The UMTS Long Term Evolution From Theory to Practice // John Wiley & Sons, 2011. ISBN 978-0-470-66025-6.
4. H. Holma and A. Toskala. LTE for UMTS - OFDMA and SC-FDMA Based Radio Access // John Wiley & Sons, first ed., 2009. ISBN 978-0-470-99401-6.
5. Jakes W. C. Microwave mobile communications. IEEE: Piscataway, NJ, 1974. 645 с.. ISBN 0-7803-1069-1.
6. Фокин Г. А. Моделирование многолучевого радиоканала // Информационные технологии и телекоммуникации. 2021. Том 9 №1. С. 59-78.
7. Миньковский М. Г. Сети Trackside в метро мы умеем строить лучше всех в мире // Первая миля. 2022. № 1. С. 36–40
8. Карасёв С. А. WI-FI в метро: архитектура сети и подземные камни // Хабр. [Электронный ресурс]. URL: <https://habr.com/ru/companies/maximatelecom/articles/332538/>
9. Feng Bao, Haoyang Yu, Hao Wang. TSN-Based Backbone Network of Train Control Management System // Wireless Communications and Mobile Computing. 2022. № 1. С. 1–12.

ОПРЕДЕЛЕНИЕ РАБОЧЕГО РЕЖИМА SNSPD ДЛЯ ДЕТЕКТИРОВАНИЯ СПУТНИКОВОГО ТЕЛЕКОММУНИКАЦИОННОГО ОПТИЧЕСКОГО СИГНАЛА

*Себряков Д.О.¹, Платонов Д.Д.¹,
Лазарев М.А.¹, Потапов Д.С.²*

*Национальный исследовательский университет
«Высшая школа экономики»,
¹департамент электронной инженерии,
²департамент прикладной математики
МИЭМ НИУ ВШЭ*

Аннотация

В настоящее время на орбиту земли выведено множество космических аппаратов, имеющих на борту в качестве полезной нагрузки терминалы лазерной связи [1, 2, 3]. Стандартные методы детектирования спутникового оптического телекоммуникационного сигнала заключаются в использовании полупроводниковых фотодетекторов в составе приемной наземной станции [4, 5, 6]. Однако полупроводниковые фотодетекторы менее чувствительны к малым оптическим мощностям приемного сигнала по сравнению со сверхпроводниковыми однофотонными детекторами [7]. Данная работа направлена на исследование возможности использования сверхпроводникового однофотонного детектора (SNSPD) в качестве приемника классического спутникового телекоммуникационного оптического сигнала. В работе приводятся результаты измерений оптического сигнала SNSPD и определяются оптимальные параметры для его работы с данным сигналом.

Введение

Терминалы лазерной спутниковой связи на малых космических аппаратах (МКА) чаще всего представляют из себя драйверы лазерных диодов с возможностью амплитудной и фазовой модуляции на разных частотах. На доступном нам МКА имеется возможность посылать на наземную станцию оптические импульсы длительностью 5 нс с частотой 25 МГц. Вследствие ослабления оптического сигнала в атмосфере, ожидаемый детектируемый сигнал будет находиться в диапазоне 150 нВт - 1.7 мкВт оптической мощности в идеальных погодных условиях.

Погодные условия могут сильно влиять на величину оптического сигнала, доходящего до наземной станции. Использование SNSPD вместо полупроводникового детектора призвано увеличить динамический диапазон приемника. Так как сверхпроводниковый детектор обладает более высокой чувствительностью по сравнению с полупроводниковым и имеет меньшее количество шумов.

Обратной стороной возможности детектирования оптических сигналов малой мощности является ослепление детектора при подаче на него оптической мощности, превышающей допустимый рабочий диапазон. Таким образом для использования SNSPD в качестве приемника спутникового телекоммуникационного оптического сигнала необходимо определить такие параметры рабочего режима детектора как: максимальная оптическая мощность регистрируемого сигнала, рабочий ток смещения детектора и значение уровня компаратора.

Результатом данной работы является определение рабочего режима детектирования SNSPD для приема спутникового оптического телекоммуникационного сигнала.

В первом разделе описан ход эксперимента, представлена схема экспериментального стенда и результаты, полученные в ходе проведения экспериментов.

Во втором разделе представлен анализ полученных результатов и определены оптимальные значения тока смещения и уровня компаратора при регистрации сигнала в диапазоне 1,7 пВт - 1.7 мкВт оптической мощности.

Ход эксперимента

Для проведения исследования использовалась оптическая схема, представленная на рисунке 1, состоящая из источника лазерного излучения (макет спутника) [3], оптоволоконного аттенюатора и сверхпроводникового однофотонного детектора с блоком электроники для счета числа импульсов.



Рис. 1. Оптическая схема экспериментального стенда. 1 - Источник лазерного излучения с длиной волны 850 нм длительностью 5 нс и частотой повторений 25 МГц; 2 - Оптоволоконный аттенюатор; 3 - Сверхпроводниковый однофотонный детектор(SNSPD); 4 - Счетчик импульсов

Оптическая мощность лазера была измерена с помощью измерителя мощности, далее эта мощность уменьшалась с помощью калиброванного аттенюатора с шагом в 5 дБ.

Результаты эксперимента

Критерием выбора значений рабочего тока смещения и уровня компаратора является среднее число отсчетов детектора.

Вероятность детектирования всех 25 млн импульсов в секунду зависит как от рабочего тока смещения детектора (от 4 до 14 мкА) так и от числа фотонов в импульсе. Так при огромном числе фотонов в импульсе вероятность детектирования всех 25 млн импульсов равна практически 100%. С такими оптическими мощностями детектор может работать только при малых рабочих токах смещения. Однако уменьшая число фотонов в импульсе в конечном итоге детектировать все 25 млн импульсов не получится даже при увеличении рабочего тока смещения.

Заключение

Использование сверхпроводникового однофотонного детектора для приема спутникового телекоммуникационного оптического сигнала является перспективным вследствие наличия возможности детектирования менее одного фотона в импульсе с высокой квантовой эффективностью. Наличие такой возможности критически важно для построения устойчивой системы всепогодной сверхскоростной спутниковой лазерной связи.

Определен рабочий режим детектирования сверхпроводникового однофотонного детектора для приема спутникового лазерного оптического телекоммуникационного сигнала на длине волны 850 нм с частотой импульсов 25 МГц. Это позволит разработать систему автоматической подстройки входной мощности оптического излучения для сверхпроводникового однофотонного детектора в космических системах связи «Спутник - Земля». Данная система позволит поддерживать на входе в детектор стабильную оптическую мощность.

Список литературы

1. C. Schieler, B. Robinson, O. Guldner, B. Bilyeu, A. Garg, K. Riesing, J.Chang, F. Hakimi, J. Brown, F. Khatri, S. Constantine, J. Burnside, D. Boroson. NASA's Terabyte Infrared Delivery (TBIRD) Program: Large-Volume Data Transfer from LEO. //33rd Annual AIAA/USU Conference on Small Satellites. -2019
2. D. Mayer, C. Kerri. CubeSat Laser Infrared Crosslink. //Document ID 20180006687. -August 4, -2018
3. A.Miller, L. Pismeniuk, A. Duplinsky, V. Merzlinkin, A. Plukchi, K. Tikhonova, I. Nesterov, D. Sevryukov, S. Levashov, V. Fetisov, S. Krasnopejev, R.Bakhshaliev. Vector—towards quantum key distribution with small satellites. //EPJ Quantum Technology. -2023
4. T. Wang, X. Zhao, Y. Song, J. Wang, X. Yu, Y. Zhang. Atmospheric Laser Communication Technology Based on Detector Gain Factor Regulation Control. 142

//IEEE Access. DOI: 10.1109/ACCESS.2021.3061504. - February 23, 2021.

5. K. Wilson, D. Antsos, L. Roberts Jr., S. Piazzolla, L. Clare, A. Croonquist. Development of the Optical Communications Telescope Laboratory: A Laser Communications Relay Demonstration Ground Station. // International Conference on Space Optical Systems and Applications (ICSOS). -2012

6. K. Schreiber, I. Prochazka, P. Lauber, U. Hugentobler, W. Schafer, L. Cacciapuoti. Ground-based demonstration of the European Laser Timing (ELT) experiment. //IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. DOI: 10.1109/TUFFC.2010.1471. -March 2010

7. Официальный сайт компании "SCONTEL" [Электронный ресурс] URL: <https://www.scontel.ru/>

ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ ПОЛЯРНЫХ И LDPC КОДОВ, ИСПОЛЬЗУЕМЫХ В СЦЕНАРИЯХ ВСЖМ

*Портной С.Л., Никитин С.Е.,
Волошин А.Д., Антошкин Г.Д.*

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В данной статье рассматриваются реализации полярных и LDPC (Low-Density Parity-Check) кодов в стандартах 3GPP 38.212 и IEEE 802.11n, применяемых в сетях FRMCS и TSN. Данная работа выполнена в рамках выполнения 1-го этапа гранта НИУ ВШЭ «Исследование методов повышения помехозащитности систем мобильной связи 4G/5G/6G, Wi-Fi, IoT с помощью сигнально-кодовых конструкций для сложных каналов в соответствии со Стратегией Развития Телекоммуникаций 2035».

Введение

С активным развитием железнодорожных магистралей и растущим объемом передаваемых данных, старые сети GSM-R уступают место новым технологиям FRMCS и TSN [1], включающим в себя компоненты стандартов 5G NR [2] и беспроводных сетей Wi-Fi на базе стандарта IEEE 802.11x.

Для сетей Wi-Fi в качестве схем кодирования выбраны сверточные и LDPC коды. Некоторое время использование LDPC-кодов в семействе стандартов 802.11 было необязательным и носило рекомендательный характер, но начиная с релиза 802.11n они стали основным классом корректирующих кодов, используемых в сетях беспроводного доступа WLAN.

В стандарте сетей мобильной связи пятого поколения (5G NR) LDPC также используются для канала данных, а полярные коды – для канала управления. Полярные коды были выбраны для сервисного канала в связи с высокой энергетической эффективностью

стью на малых и средних длинах кодовых блоков и относительной простоте реализации.

LDPC-коды

В 1960 году Роберт Галлагер предложил класс линейных блочных кодов с низкой плотностью проверок на четность (LDPC-коды) [3], с высокой эффективностью, близкой к границе Шеннона.

Кодирование LDPC-кодов осуществляется посредством матричного умножения порождающей матрицы на последовательность информационных битов. В настоящий момент, на практике используются циклические и квазициклические LDPC-коды, что позволяет сократить потребление ресурсов при кодировании.

Основным способом получения порождающей матрицы квазициклического LDPC-кода является преобразование базовой матрицы с помощью алгоритма расширения, который, увеличивает матрицу в Z раз и делает ее разреженной. Базовые матрицы могут быть получены различными способами, такими как случайная генерация с эмпирической проверкой полученного кода и построение с помощью графа Таннера.

Рис.1. Порождающая и базовая матрица LDPC-кода

Порождающая матрица состоит из двух частей: проверочной (parity) и единичной (identity):

$$G_{K \times N} = [I_{K \times K} \ P_{K \times (N-K)}]$$

Для получения проверочной матрицы необходимо взять транспонированную проверочную часть порождающей матрицы и дописать справа единичную матрицу:

$$H_{(N-K) \times N} = [P_{(N-K) \times K}^T \ I_{(N-K) \times (N-K)}]$$

Декодирование LDPC

Декодирование LDPC кодов [4] сводится к выполнению условия, что произведение проверочной матрицы на принятую декодером последовательность символов равняется нулю:

$$H * X = 0$$

где H – проверочная матрица, X – принятое сообщение.

Для декодирования LDPC-кодов существуют три базовых алгоритма: декодирование с жестким реше-

нием (Bit-flipping), декодирование по вероятностям (Sum-Product) и декодирование по вероятностям с аппроксимацией (Min-Sum). Все эти алгоритмы являются итеративными.

Алгоритм с жестким решением основан на инверсии битов и включает в себя вычисление проверок на четность, инверсию бита, участвовавшего в большинстве невыполненных проверок, и проверку полученной последовательности на принадлежность к кодовым словам.

Алгоритмы с мягким решением, такие как декодирование по вероятностям и декодирование по вероятностям с аппроксимацией, используют вектор с вещественными значениями и логарифмические отношения правдоподобия для получения информации о канале и имеют куда большую точность, предлагая повышение энергетической эффективности до 3 дБ, по сравнению с жестким решением. На каждой итерации этих алгоритмов вычисляются сообщения "бит-проверка" и "проверка-бит", формируется вектор апостериорной вероятности и проверяется его принадлежность кодовым словам.

LDPC-коды в 5G NR

Цепочка кодирования NR LDPC начинается с сегментации кодовых блоков. Затем к сегментированным блокам присоединяется циклический избыточный код (CRC). Далее следует кодирование LDPC, где кодер обрабатывает блоки с использованием квазициклической структуры. Для повышения производительности процесс кодирования включает в себя перемежитель с систематическим приоритетом битов, с числом строк, равным порядку квадратурной амплитудной модуляции (QAM).

В стандарте 5G NR существует два класса базовых порождающих матриц для LDPC-кода: BG1 и BG2 с размерами 46 x 68 и 42 x 52 соответственно. Эти матрицы имеют следующую структуру:

$$H = \begin{bmatrix} A & E & O \\ B & C & I \end{bmatrix}$$

Для BG1: A: 4 x 22; E: 4 x 4; O: нулевая матрица 4 x 42; B: 42 x 22; C: 42 x 4; I: единичная матрица 42 x 42.

Для BG2: A: 4 x 10; E: 4 x 4; O: нулевая матрица 4 x 38; B: 38 x 10; C: 38 x 4; I: единичная матрица 38 x 38

Для каждого элемента базовой матрицы выполняются следующие действия:

- Если элемент равен нулю, то он заменяется на нулевую матрицу размером $Z_C * Z_C$;
- Если элемент равен единице, то он заменяется на матрицу перестановки $I(P_{i,j})$ размером $Z_C * Z_C$, где матрица получена из единичной матрицы циклическим сдвигом вправо $P_{i,j}$ раз;
- $P_{i,j}$ указаны в спецификации стандарта 5G NR;
- Z_C выбирается в зависимости от длины кодового блока таким образом, что $K_B * Z_C \geq K'$ (где K_B – базовая длина, K' – длина кодового блока);

Если длина сообщения B больше, чем максимальная длина кодового блока, то производится сегментация входной последовательности битов и к каждому полученному блоку добавляется дополнительная проверка избыточности L длиной 24 бита. Максимальная длина кодового блока равна 8448 бит для BG1 и 3840 бит для BG2. Для BG1 базовая длина равна 22, а для BG2 базовая длина равна 10 при $B > 640$, 9 при $560 < B \leq 640$, 8 при $192 < B \leq 560$ и 6 в остальных случаях.

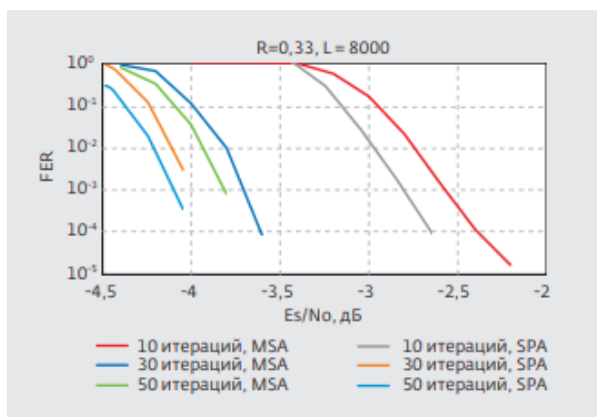


Рис.2. Оценка эффективности LDPC-кода

Согласно оценке эффективности, приведенной на рис.2, Sum-Product алгоритм является более производительным, чем Min-Sum, но, поскольку он требует больших вычислительных мощностей, в 5G NR применяется Min-Sum.

LDPC-коды в 802.11n

В стандарте Wi-Fi 802.11n [4] используются квазициклические (QC) LDPC-коды с фиксированной длиной блока $L = 672$ бита и скоростями кода $R = \frac{1}{2}, \frac{3}{4}, \frac{5}{8}$ и $\frac{13}{16}$.

Процесс кодирования помещает биты сообщения в целое число символов OFDM и целое число кодовых слов LDPC, используя следующие шаги:

1. Определяется минимальное количество символов OFDM на основе длины сообщения и количества бит на поднесущую (SC).
2. Определяется количество кодовых слов и их длина, согласно таблице, описанной в стандарте.
3. Определяется количество сокращаемых битов, где сокращение определяется как процедура вставки нулей после битов сообщения перед вычислением битов четности.
4. Вычисляются биты четности для каждого кодового слова, используя матрицы проверки на четность. Биты сокращения удаляются из каждого кодового слова после вычисления битов четности.
5. Операция выкалывания выполняется если количество закодированных битов на 159 больше, чем общее количество закодированных битов, которые помещаются во все символы OFDM.

В противном случае, выполняется повторение закодированных битов для заполнения всех битов в OFDM-символах.

6. Выполняется объединение кодовых слов, демultipлексирование битов кодового слова в пространственные потоки (stream parsing) и закодированные биты модулируются в OFDM символы.

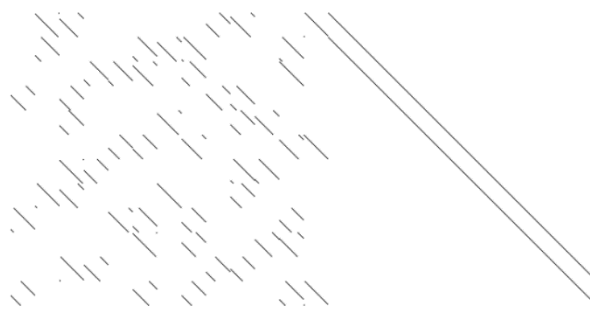


Рис.3. Структура проверочной матрицы Wi-Fi LDPC для скорости кода 1/2

Проверочные матрицы для остальных скоростей кода имеют структуру, аналогичную кодовой скорости $R = \frac{1}{2}$, изображенную на рис.3.

Полярные коды

Полярные коды предложенные в 2009 году [5], являются классом линейных блочных кодов, которые представляют собой способ кодирования информации, основанный на трансформации исходного канала, позволяющей вплотную приблизиться к пределу Шеннона.

Принцип построения полярных кодов основан на поляризации канала – операции преобразования канала связи в N его независимых копий, в которых вероятность ошибки при передаче данных стремится к 0 (информационные подканалы) или к 1 (замороженные подканалы). В результате преобразований, надежные биты становятся информационными символами, а менее надежные приобретают фиксированные значения.

Процесс декодирования полярных кодов в 5G NR

Алгоритм последовательной отмены (Successive Cancellation - SC), являющийся фундаментом для более продвинутых версий, основан на рекурсивном обходе дерева и принятия решения по каждому биту, его сложность описывается выражением $O(N * \log N)$.

Одним из первых способов повысить эффективность стал алгоритм SCL (Successive Cancellation List), основным отличием которого является хранение списка возможных кодовых слов. Его основным недостатком по сравнению с SC является необходимость хранить полученные варианты в списке длиной L , что дает сложность алгоритма $O(LN * \log N)$.

Одним из слабых мест SCL является низкая эффективность при декодировании длинных последовательностей. Для решения этой проблемы в 2016 году был предложен алгоритм SSCL (Simplified Successive Cancellation List) [6], основным преимуществом которого является идентификация составляю-

ших кодов в структуре полярного кода, которые могут быть декодированы независимо. Это позволяет сократить количество необходимых вычислений без потери качества исправления ошибок.

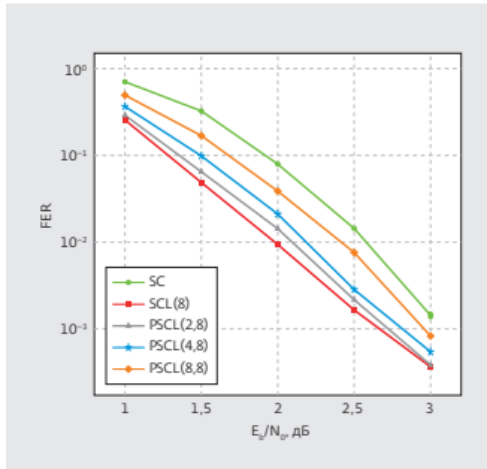


Рис.4. Оценка эффективности алгоритма PSCL

В 2017 году был представлен алгоритм списочного декодера PSCL (Partitioned Successive Cancellation List) [7]. Суть этого алгоритма заключается в разделении большого дерева и декодировании каждого из получившихся малых деревьев с помощью списочного декодера. Данный алгоритм позволяет существенно оптимизировать использование памяти ценой ухудшения качества коррекции ошибок. Однако этот недостаток можно исправить за счет использования CRC. На рис.4 представлен график, демонстрирующий неэффективность PSCL без CRC по сравнению с SCL. Отношение сигнал-шум равно 2 дБ, длина кодового слова 1024 бита, длина сообщения 512 бит.

На рис.5 приведены результаты моделирования эффективности алгоритмов SCL и PSCL с CRC по сравнению с LDPC WiMAX.

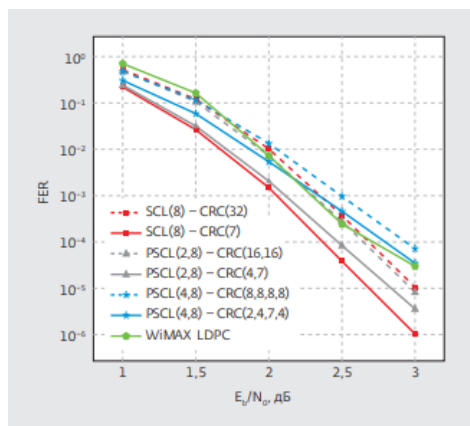


Рис.5. Оценка эффективности PSCL с добавлением CRC

Заключение

LDPC коды и полярные коды стандартов 5G и Wi-Fi обладают высокой производительностью и

относительно низкой сложностью декодирования, что позволяет добиться низкой вероятности ошибки на блок даже в условиях сложных каналов, таких как канал поезд-земля в условиях ВСЖМ.

Для LDPC кодов оптимальным для практического применения является декодер Min-Sum с 20-30 итерациями, в то время как для полярных кодов предпочтительным декодером является PSCL.

Список литературы

1. Портной С.Л., Никитин С.Е., Клюев Н.С., Сахаутдинов Ш.Р. Обзор методов реализации подвижной железнодорожной связи // Первая миля. 2023. №8. С. 26–35.
2. 3GPP TS 38.212. Multiplexing and channel coding (Release 18). 2023.
3. Gallager R. Low-density parity-check codes // IRE Transactions on Information Theory. 1962. Vol. 8. No 1. PP. 21–28.
4. IEEE Std 802.11n-2009
5. Arıkan E. Channel polarization: A method for constructing capacity achieving codes for symmetric binary-input memoryless channels // IEEE Transactions on Information Theory. 2009. Vol. 55. No. 7. PP. 3051–3073.
6. Hashemi S.A. et al // Simplified Successive Cancellation List decoding of polar codes // International Symposium on Information Theory (ISIT). 2016. PP. 815–819
7. Hashemi S.A. et al. Partitioned List Decoding of Polar Codes: Analysis and Improvement of Finite Length Performance // GLOBECOM 2017 IEEE Global Communications Conference, Singapore. 2017. PP. 1–7

ПРЕДМЕТНО-ОРИЕНТИРОВАННЫЙ ЯЗЫК ДЛЯ ОПТИМИЗАЦИИ СИСТЕМ ИНТЕРНЕТА ВЕЩЕЙ

Эбрахим А.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В этой работе предлагается предметно-ориентированный язык, направленный на упрощение описания структур систем Интернета вещей, формулирования требований и оптимизации проектирования этих систем. Язык позволяет генерировать код Python, который с помощью методов многокритериальной оптимизации выбирает и распределяет ресурсы в системе, что приводит к повышению эффективности систем Интернета вещей.

Введение

Интернет вещей (англ Internet of Things или IoT) — это сеть физических устройств, оснащенных датчиками и исполнительными механизмами, которые могут подключаться, обмениваться и хранить дан-

ные. Данная технология повышает производительность и качество жизни. Однако точное и оптимальное проектирование и планирование систем Интернета вещей имеет решающее значение. Инструменты вычислительной оптимизации могут помочь, но этот процесс может быть громоздким и трудоемким, так как не все являются экспертами в данной области. Предметно-ориентированные языки (англ. Domain-Specific Languages или DSL) могут упростить эту задачу, предоставляя высокую абстракцию для сложных систем. В этой работе представлен новый предметно-ориентированный язык под названием OptIoT, предназначенный для решения проблем оптимизации инфраструктуры в области IoT.

Обзор литературы

Различные предметно-ориентированные языки были разработаны для ускорения разработки, упрощения сложности и удовлетворения разнообразия. ThingML предлагает абстрактный способ описания вещей и взаимодействий, который можно компилировать в такие языки, как Java, JavaScript или C [1]. На основе языка SoaML был создан SoaML4IoT для проектирования сервис-ориентированных систем Интернета вещей [2]. IoTECS упрощает создание симуляторов систем от границы сети до облака [3], DSL-4-IoT — это язык визуального моделирования на основе UML [4]. Node-RED упрощает связывание компонентов, API и веб-сервисов, используя подход программирования на основе потоков (англ. Flow-based programming) [5]. MOCSL упрощает разработку собственных приложений для интеллектуальных объектов [6], Monitor-IoT помогает с мониторингом систем Интернета Вещей [7]. IRON использует правила ECA (англ. Event-Condition-Action) для программирования систем Интернета вещей [8].

Эти языки ориентированы на генерацию кода и моделирования, но в этой работе был применен OptIoT для оптимизации инфраструктуры Интернета вещей путем выбора сетевых устройств и компонентов для повышения производительности, экономической эффективности или обеспечения баланса между обоими факторами.

Описание технического решения

OptIoT- новый фреймворк, который позволяет пользователям определить свою проблему проектирования систем Интернета Вещей в стандартизованном формате. Определяя спецификации ресурсов, архитектуру системы и требования в простом синтаксисе OptIoT, пользователи могут использовать мощь rumpoo [9], библиотеки Python, созданной для решения задач многокритериальной оптимизации. Можно сказать, что основная задача, выполняемая OptIoT, — преобразовать требования в форму, используемую rumpoo для запуска процесса оптимизации, что математически можно выразить как:

$$\min F_a(x) \quad a = 1, \dots, A \quad (1)$$

$$f_b(x) = 0 \quad b = 1, \dots, B \quad (2)$$

$$f_c(x) \leq 0 \quad c = 1, \dots, B \quad (3)$$

$$x_n^L \leq x_n \leq x_n^U \quad (4)$$

где F_a — целевая функция, (2) определяет ограничения равенства, (3) определяет ограничения неравенства, а (4) определяет верхние и нижние пределы переменных решения.

textX был выбран как платформа для создания OptIoT благодаря ее совместимости с Python и с огромным набором его библиотек, предназначенных для научных целей. textX специально создан для проектирования предметно-ориентированных языков с использованием парсера Areggio PEG (англ. Parsing Expression Grammar). Используя грамматику, textX генерирует как парсер, так и метамодель, интегрируя классы Python для построения объектного графа. Структура и рабочий процесс OptIoT изображены на Рис.1.

Синтаксис OptIoT основан на простых элементах кода, которые в дополнение к правилам можно разделить на фундаментальные, архитектурные, физические, географические элементы и элементы оптимизации. Элементы кода абстрактны и просты в использовании для неопытного пользователя.

Результаты

В тестовом сценарии, OptIoT использовался для оптимизации размещения и количества маршрутизаторов в промышленном здании. В установке использовались умные беспроводные устройства на основе плат ESP32, собирающие данные с различных датчиков. Система требовала, чтобы устройства были подключены к серверу через маршрутизаторы.

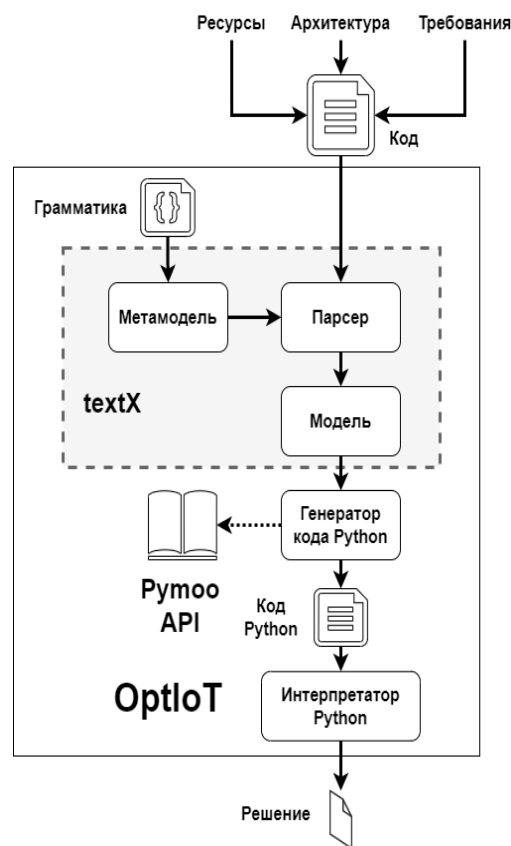


Рис.1. Структура и рабочий процесс OptIoT

OptIoT анализирует входные данные, такие как спецификации маршрутизаторов, планировка здания, потери сигнала через стены, а также программу, определяющую структуру и местоположения устройств. Для прогнозирования потерь сигнала внутри помещений использовалась модель Мотли-Кинана [10], учитывая такие факторы, как материал стен и расстояние между передатчиком и приемником. Как метод оптимизации, использовался метод роя частиц [11]. Поскольку метод оптимизации может дать несколько оптимальных решений на фронте Парето, выбиралось решение с минимальной стоимостью. Предложенное решение было реализовано и оказалось работоспособным.

Заключение

В этой работе представлен предметно-ориентированный язык под названием OptIoT, предназначенный для оптимизации инфраструктуры систем Интернета вещей. Язык объединяет системные требования и ресурсы в задачу оптимизации с использованием алгоритмов многокритериальной оптимизации из библиотеки ruoo. Протестированный в практическом сценарии, этот язык показал свою эффективность. Хотя он не обеспечивает генерацию кода для узлов Интернета вещей, дальнейшие исследования могут привести к созданию более универсального инструмента для быстрого и оптимизированного проектирования.

Список литературы

1. Github site. 2023. - URL: <https://github.com/TelluIoT/ThingML> (date of treatment: 26.12.2023).
2. Costa, B. Modeling SOA-based IoT applications with SoaML4IoT / B. Costa, P. Pires, F. Delicato // Proc. IEEE 5th World Forum Internet Things (WF-IoT). 2019. - P. 496-501.
3. Li, J. Domain-specific language for simulation-based testing of IOT edge-to-cloud solutions / J. Li, S. Nejati, M. Sabetzadeh, M. A. McCallen // Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems. 2022.
4. Eterovic, T. An Internet of Things visual domain specific modeling language based on UML / T. Eterovic, E. Kaljic, D. Donko, A. Salihbegovic, S. Ribic // Proc. XXV Int. Conf. Inf. Commun. Autom. Technol. (ICAT). 2015. -P. 1-5.
5. Node-RED site. 2023. - URL: <https://nodered.org/> (date of treatment: 30.12.2023).
6. García, C.G. Midgar: Creation of a graphic domain-specific language to generate smart objects for Internet of Things scenarios using model-driven engineering / C. G. García, D. Meana-Llorián, García- V. Díaz, A. C. Jiménez, J. P. Anzola // IEEE Access. 2020. - P. 141872-141894.
7. Erazo-Garzón, L., A Domain-Specific Language for Modeling IoT System Architectures That Support Monitoring / L. Erazo-Garzón, P. Cedillo, G. Rossi, J. Moyano // IEEE Access. 2022. -P. 61639-61665.
8. Cacciagrano, D.R. IRON: Reliable domain specific language for programming IoT devices / D.R.

Cacciagrano, R. Culmone // Internet of Things. 2020. - P. 100020.

9. Blank, J. Pymoo: Multi-Objective Optimization in Python / J. Blank, K. Deb. IEEE Access. 2020. -P. 89497-89509.

10. Kar, K. Motley Keenan model of in-building coverage analysis of IEEE 802.11n WLAN signal in electronics and communication engineering department of National Institute of Technology Durgapur / K. Kar, S. Datta, M. Pal, R. Ghatak // 2016 International Conference on Microelectronics, Computing and Communications (MicroCom). - Durgapur, 2016. -P. 1-6.

11. Kennedy, J. Particle swarm optimization / J. Kennedy and R. Eberhart. Proceedings of ICNN'95 - International Conference on Neural Networks. 1995. P. 1942-1948.

ПРИМЕНЕНИЕ НЕЙРОННОЙ СЕТИ ДЛЯ ПРОГНОЗИРОВАНИЯ ТОЧКИ НАСЫЩЕНИЯ В СЕТЯХ НА КРИСТАЛЛЕ

*Маркова Э.С, Макагонова В.Д, Нигматуллин Н.Р.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В данной работе представлено исследование высокоуровневых моделей, в частности был создан классификатор высокоуровневых моделей и была проведена их классификация. На основе проведенных исследований высокоуровневых моделей было предложено исследовать нейронные сети для более быстрого поиска одного из выходных параметров моделирования: точки насыщения.

Введение

На сегодняшний день задачи, которые должны решаться с помощью систем на кристалле (СнК), становятся более сложными и трудоемкими, поэтому необходимо применять многопроцессорные СнК, в кодсистему связи сетей на кристалле (СтнК) так, чтобы она удовлетворяла все возрастающим требованиям.

Система на кристалле (СнК, англ. System-on-a-Chip, SoC) [1] — электронная схема, выполняющая функции целого устройства (например, компьютера) и размещенная на одной интегральной схеме.

Сеть на кристалле (СтнК, англ. Network-on-a-Chip, NoC) [2] — это сеть - на основе подсистемы связи на интегральной схеме («микрочип»), чаще всего между модулями в СнК. Это сеть маршрутизаторов с коммутацией пакетов между модулями СнК.

Классификация высокоуровневых моделей СтнК

В процессе проектирования СтнК используются высокоуровневое моделирование. Данный процесс осуществляется с помощью высокоуровневых имитационных моделей.

тационных моделей. Высокоуровневая имитационная модель представляет собой набор сущностей (маршрутизаторов, роутеров, пакетов и т.д.), а также набор правил, по которым данные сущности взаимодействуют друг с другом.

Существует большое количество высокоуровневых моделей, поэтому для облегчения процесса проектирования необходима их классификация по различным признакам.

При проведении исследования были выявлены следующие варианты классификации:

1. Функциональная классификация.

Все симуляторы СтнК в зависимости от цели их применения сети общего назначения, которые предназначены для моделирования любых сетей в целом, но могут быть использованы и для СтнК, и симуляторы, разработанные специально для моделирования СтнК.

2. Языки программирования

Высокоуровневые СтнК обычно разрабатываются на таких языках программирования как C/C++ (и его производных), Python и Java. Выбор метода программирования зависит от цели и задач, которые стоят перед исследователем.

3. Модифицируемость моделей

Некоторые модели имеют открытый исходный код, что дает возможность их модификации под определенные цели и проведение исследований по проектированию СтнК с заданными характеристиками.

4. Параметры сети.

Поддерживаемые топологии СтнК, методы коммутации, архитектура маршрутизаторов, тип трафика, алгоритмы маршрутизации, алгоритмы арбитража.

Опираясь на данные элементы классификации, было создано классификационное дерево (рис.1).



Рис.1. Классификационное дерево для высокоуровневых моделей СтнК

Применение классификатора для высокоуровневых моделей

Воспользовавшись полученным деревом классификации, была проведена классификация. Для рассмотрения в исследование были использованы следующие модели: Toraz, Noxim, BookSim, NoC DES Simulator [3], HORNET, VNOC 2.0, NNSE, PYOCN и Nigram.

На примере модели BookSim 2.0 [4] покажем использование дерева классификации. Данный симулятор используем арбитражный алгоритм round robin. Архитектура маршрутизатора – input-queued. Поддерживаемые топологии: Single node, Mesh, Concentrated mesh, Torus, Butterfly, Flattened butterfly, Quad tree, Concentrated multi-dimensional octagon, Fat tree, MECS, Dragonfly, Kncube. Алгоритмы маршрутизации: XY, Adaptive XY, Random adaptive, Minimal adaptive, Planar adaptive, Nearest common ancestor, Adaptive nearest common ancestor, Chaos, Dimension order. Типы трафика: Random uniform, Bit-reverse, Bit-complement, Bit-reverse, Shuffle, Transpose, Tornado, Random permutation. Симулятор был разработан на высокоуровневом языке программирования C++ для UNIX-образных систем. Симулятор имеет открытый исходный код, расположенный в репозитории github [5], и документацию в открытом доступе, что позволяет свободно модифицировать модель. Данная модель специализирована для моделирования сетей на кристалле, что позволяет проводить более точное моделирование с использованием большего числа параметров.

СППР (Система поддержки принятия решений)

При большом количестве узлов высокоуровневое моделирование становится трудозатраты и длительным процессом, поэтому возникает необходимость в создании инструментов для снижения трудозатрат и времени на моделирование.

Для уменьшения трудозатрат на этапе высокоуровневого моделирования СтнК разработан модуль СППР. Модуль СППР, представляет собой отдельный архитектурный блок., входящий в САПР высокоуровневого моделирования СтнК.

После ввода пользователем задания, системой производится валидация входных данных и оценка задания системой поддержки принятия решений (СППР). СППР позволяет оценить производилось ли моделирование СтнК данной конфигурации ранее. Если проводилось, то система передает результаты моделирования из базы данных пользователю, если нет, то система помогает пользователю выбрать необходимые методы и высокоуровневые модели для проведения моделирования.

В ходе проекта было проведено исследование для упрощения использования САПР, было предложено провести исследование нейронных сетей и высокоуровневых моделей для более быстрого поиска точки насыщения. Точка насыщения находится при помощи двух параметров: оптимальной скорости генерации данных и оптимального значения пропускной способности. Но главной проблемой при поиске

точки насыщения является время моделирования, в результате чего было предложено использовать нейронную сеть для предсказания двух вышеупомянутых параметров для более быстрого поиска данной точки. Для создания датасета для обучения нейронной сети использовалась высокоуровневая модель booksim с топологией mesh. Ниже приведен обзор нейронных сетей для предсказания числовых рядов.

Искусственные нейронные сети (ИНС) являются интеллектуальными моделями биологических нейронов и эффективно используются для прогнозирования числовых рядов [6]. Они обладают способностью к обучению [7] и имеют нелинейную связь между входом и выходом для аппроксимации любой непрерывной функции. Рассмотрим некоторые нейронные сети.

Feedforward Neural Networks (FNN) — это сеть, которая не имеет рекуррентных связей, что означает, что сигналы могут проходить только в одном направлении. Данные передаются от входного слоя для дальнейшей обработки.

Рекуррентные нейронные сети имеют итерационные циклы в своей архитектуре и получают выход, используя обратную связь от своих блоков [8]. Они являются одним из наиболее мощных инструментов для прогнозирования временных рядов, позволяя моделировать зависимости между последовательными значениями временного ряда и запоминать информацию о прошлых значениях входных данных.

LSTM (Long Short-Term Memory) — это рекуррентная нейронная сеть, которая может сохранять информацию в течение длительного периода времени, что делает ее особенно полезной в задачах обработки естественного языка, обработки звука и распознавания речи.

Jordan neural network — это сеть обратной связи, в которой выход сети используется как вход на предыдущих временных шагах. Такая сеть может прогнозировать последовательности данных, особенно для временных рядов.

Сверточная нейронная сеть (Convolutional Neural Network, CNN) — это тип нейронной сети, который широко используется для классификации изображений, а также для обработки и анализа других типов данных, таких как звуковые сигналы и текстовые данные.

Выбор нейронной сети

Согласно приведенному выше обзору, была выбрана нейронная сеть JNN (Jordan Neural Network). При написании кода использовалась библиотека Keras языка программирования Python, а конкретно специальная функция SimpleRNN.

Сначала мы загружаем набор данных. Данный датасет состоит из 7 столбцов: 5 входных и 2 выходных набора данных. На рис.2 изображены названия столбцов. Первые два столбца: n и m, являются аргументами топологии mesh, которая использовалась при моделировании высокоуровневых модели bookism. Routing function представляет собой название функции маршрутизации использованной во время моделирования (dim_order и dor), т.к. нейрон-

ная сеть не может использовать данные типа string, данные в столбце были заменены на 1 для dim_order и 0 для dor. Следующие два столбца Virtual channels number и Virtual channel buffer size или количество и размер виртуальных каналов. Данные в этих столбцах варьировались от 1 до 4. И последние два столбца Optimal data generation rate и Optimal throughput value или оптимальная скорость генерации данных и оптимальное значение пропускной способности, данные из этих столбцов используются для поиска точки насыщения, которая и является целью использования нейронной сети.

| n | m | Routing function | Virtual channels number | Virtual channel buffer size | Optimal data generation rate | Optimal throughput value |
|---|---|------------------|-------------------------|-----------------------------|------------------------------|--------------------------|
| 0 | 2 | 1 | 1 | 1 | 0.30 | 0.296800 |
| 1 | 2 | 1 | 1 | 2 | 0.50 | 0.513934 |
| 2 | 2 | 1 | 1 | 3 | 0.90 | 0.820686 |
| 3 | 2 | 1 | 1 | 4 | 1.10 | 1.053134 |
| 4 | 2 | 1 | 2 | 1 | 0.30 | 0.286534 |

Рис.2. Датасет

После загрузки датасета определяем количество шагов, которые модель будет использовать для прогнозирования (lookback). Затем мы преобразуем данные с помощью MinMaxScaler и создаем входные и выходные данные для обучения модели. Мы разделяем данные на обучающий и тестовый наборы и строим модель RNN с одним слоем и 3000 нейронами, а также один выходной слой с 2 нейронами. Мы используем функцию активации ReLU и линейную функцию активации для выходного слоя. Модель компилируется с помощью оптимизатора adam и функции потерь mean_squared_error. Мы обучаем модель на обучающем наборе в течение 300 эпох. Наконец, мы оцениваем модель на тестовом наборе и выводим точность прогнозирования.

Для тестовой выборки функция потерь (loss) равна 0.0011, а точность (accuracy) модели на этой выборке – 0.8493. Для обучающей выборки – функция потерь – 0.0000007, а точность равна 0.8806.

Заключение

Согласно исследованию, можно сделать следующие выводы, при создании дополнения к подпрограмме СППР рекомендуется создавать нейронные сети для каждой топологии высокоуровневой модели, т.к. хоть большая часть входных параметров одинакова, аргументы и структура самих топологий сильно отличаются, что с учетом ограничений самой высокоуровневой модели может привести к низким результатам точности прогноза. Также рекомендуется использовать нейронную сеть JNN, т.к. она показала свою эффективность при прогнозировании числовых рядов. Кроме того, в ходе исследования при помощи САПР «Высокоуровневого моделирования» была собрана база данных для различных моделей.

Список литературы

1. Система на кристалле — Википедия // Википедия URL: https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_%D0%BD%D0%B0_%D0%BA%D1%80%D0%B8%D1%81%D1%82%D0%B0%D0%BB%D0%BB%D0%B5.
2. Сеть на кристалле — Network on a chip — Википедия // Википедия URL: https://wiki5.ru/wiki/Network_on_a_chip.
3. Jones M. NoCsim: a versatile network on chip simulator. 2005.
4. Pérez I. et al. BST: A BookSim-based toolset to simulate NoCs with single-and multi-hop bypass // 2020 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS). IEEE, 2020. P. 47–57.
5. BookSim Interconnection Network Simulator // Github URL: <https://github.com/booksim/booksim2>.
6. Malik N. Artificial neural networks and their applications // arXiv preprint cs/0505019. 2005.
7. Mushtaq M.F. et al. An innovative cognitive architecture for humanoid robot // International Journal of Advanced Computer Science and Applications. Science and Information (SAI) Organization Limited, 2017. Vol. 8, № 8.
8. Martens J., Sutskever I. Learning recurrent neural networks with hessian-free optimization // Proceedings of the 28th international conference on machine learning (ICML-11). 2011. P. 1033–1040.

КОНЦЕПЦИЯ ПРОГРАММНО-АППАРАТНОЙ ПЛАТФОРМЫ ДЛЯ РЕАЛИЗАЦИИ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ В СФЕРЕ ИНТЕРНЕТА ВЕЩЕЙ: ВОЗМОЖНОСТИ, ПРИМЕНИМОСТЬ И ПЕРСПЕКТИВЫ

Еременко Е.А.¹, Лыжин И.Г.^{1,2}

*¹Национальный исследовательский университет «Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ,*

²Сколковский институт науки и технологий

Аннотация

В работе предлагается концепция образовательной платформы интернета вещей и возможности её использования при проведении практикумов на примере лабораторных работ (кейсов) IT Академии Samsung (трек Интернет вещей).

Введение

Интернет вещей играет важную и актуальную роль в современном мире, от домашней автоматизации и безопасности до медицины [1], промышленного производства и управления ресурсами городов и регионов. В 2016 году число подключенных к Интернету устройств превышало 6.5 млрд [2]; в 2023 году число устройств Интернета вещей превысило 15 миллиардов [3]. При повышении доступности микроконтроллеров и уменьшении их стоимости, они

начинают использоваться в тех сферах, где ранее такие решения были экономически нецелесообразными [4].

В связи с растущим внедрением IoT-устройств и технологий возрастает значимость вопросов образования и приобретения навыков работы со стеком технологий Интернета вещей.

Проект образовательной платформы, концепция которого предлагается в работе, ставит задачу разработки комплексной платформы для сбора, передачи, обработки, анализа и визуализации данных, полученных от различных устройств интернета вещей, в том числе образовательно-демонстрационного стенда, который может применяться на практических, лабораторных и лекционных занятиях.

В работе рассматриваются преимущества предлагаемой программно-аппаратной платформы, возможные технологии и продукты для реализации, проанализирована применимость платформы условиях очного, смешанного (гибридного) и онлайн (дистанционного) формата обучения.

Существующие решения

Существующие решения по внедрению технологии IoT в образование связаны с определенными сложностями, такими как: высокая стоимость, сложность масштабирования, проблемы безопасности и другие [5].

Близким решением является Data Hub для устройств интернета вещей на базе стека ELK (Elasticsearch, Logstash and Kibana [6]). Однако и у него есть существенные недостатки: в работе детально рассмотрена только программная часть, отсутствует адаптация для образовательных проектов, высокая ресурсоемкость, специфика использования внутреннего языка запросов QueryDSL, проблемы с безопасностью, отсутствие публичного API у Kibana, недостаточная гибкость.

Рассмотрим также проект Yandex IoT Core [7], который является одним из сервисов платформы Yandex Cloud и представляет собой облачный MQTT-сервер. Этот проект с проприетарным исходным кодом предоставляет возможность устанавливать связь между устройствами и локальными или облачными ресурсами, но при этом не даёт возможности формирования наборов данных и их визуализации. Кроме того, изначально решение не сфокусировано на применении в образовательном процессе.

Компанией МТС был разработан продукт МТС IoT HUB [8]. Платформа работает по модели PaaS (Platform as a Service, «платформа как услуга») и позволяет создавать свои решения для управления устройствами интернета вещей. Функционал платформы включает в себя поддержку протоколов MQTT, COAP и HTTP. Однако, как и Yandex IoT Core, платформа не является self-host, не сфокусирована на применении в образовании и не является решением с открытым исходным кодом.

Решением, ориентированным на применение в образовании, является платформа Arduino Education [9]. Это проект с открытым исходным кодом, который предоставляет поддержку протокола MQTT, библиотек и SDK для различных устройств, а также

ограниченную возможность формирования наборов данных. Но стоит отметить, что платформа не предусматривает возможности размещения на собственных серверах (не является self-host) и не предоставляет возможности визуализировать данные и обмениваться ими с другими пользователями.

Итак, в настоящий момент времени не существует готового и ориентированного на российский рынок расширяемого сервиса, удобного для использования в исследовательских и образовательных целях, который отвечал бы следующим требованиям: поддержка нескольких протоколов, открытый исходный код, возможность использования по модели PaaS, возможность визуализации данных и обмена ими, отсутствие необходимости регистрации для получения базового функционала (что важно, например, при проведении открытых мастер-классов, где за короткое время участники должны попробовать работать с сенсорами, а не тратить время на регистрацию).

Концепция образовательной платформы

Мы предлагаем концепцию платформы (в дальнейшем именуемую IoT Educational DataHub), предназначенной для сбора, хранения и анализа данных, полученных от различных устройств Интернета вещей (IoT). Платформа позволит пользователям получать и визуализировать информацию в режиме реального времени, анализировать данные и обмениваться ими с другими пользователями платформы.

Предлагаемая платформа предоставляет ряд преимуществ для образования и исследований. Во-первых, она позволяет собирать данные от множества устройств, что позволяет проводить комплексные исследования и эксперименты. Во-вторых, платформа обеспечивает визуализацию данных, что делает процесс анализа более понятным и наглядным. В-третьих, платформа будет поддерживать обмен данными между пользователями. В-четвертых, изначально сфокусирована на использовании в образовательном процессе и сопровождается методическими рекомендациями, что позволит преподавателям и студентам изучать принципы работы устройств интернета вещей, анализировать полученные данные, а также делиться ими.

Одной из составляющих проекта является разработка демонстрационно-образовательного стенда, который позволит наглядно продемонстрировать работу платформы и ее возможности.

Архитектура платформы

Предполагается, что IoT Educational DataHub будет включать следующие компоненты.

Устройства IoT: различные датчики, контроллеры, смартфоны или другие устройства, способные собирать данные и передавать их в систему. В качестве примера активно используются сейчас устройств можно привести Raspberry Pi, ESP32, ESP8266, Arduino Uno, Pi Pico W.

API (back-end): разделяется на открытую (необходимую для взаимодействия с IoT устройствами) и закрытую (реализующую функции добавления пользователей, удаления, функций, необходимых для front-end) части. Необходим для доступа к данным,

хранящимся в системе; кроме того, API позволит разработчикам приложений и умных устройств, в том числе другим студенческим проектам похожей тематики, интегрироваться с IoT Educational DataHub. В качестве языка программирования предлагается использовать Python 3.x с использованием фреймворка Flask.

Сервер хранения данных (база данных): этот компонент отвечает за хранение данных, полученных от устройств IoT, в структурированном и удобном для анализа виде. Для этого планируется использовать базу данных с поддержкой SQL-запросов (MariaDB), и при необходимости NoSQL решения, например, MongoDB.

Графический интерфейс (Front-end): этот компонент представляет собой веб-приложение, которое позволяет пользователям просматривать данные, полученные от устройств IoT, и непосредственно визуализирует данные, превращая их в графики и диаграммы, вывода состояния устройств. На платформе должны быть доступны различные типы графиков, включая линейные графики, столбчатые диаграммы, круговые диаграммы и другие. Кроме графиков, информацию с датчиков возможно визуализировать с помощью тепловых карт, 3D-моделей (цифровых двойников); для отдельных категорий данных удобно реализовать оценку с помощью распространенных статистических метрик (коэффициент корреляции, среднее, дисперсия и так далее). Также предусмотрены варианты визуализации данных с датчиков, включающие создание интерактивных карт, использование анимации для демонстрации изменений в данных, создание интерактивных блок-схем для отображения процессов.

Аналитический модуль: этот компонент предназначен для анализа данных. Он включает в себя инструменты статистического анализа, использование методов машинного обучения и прогнозирования (например, такие модели машинного обучения, как регрессионный анализ [10], случайный лес [11], [12], градиентный бустинг [13], [14] и так далее). Для реализации аналитического модуля также предлагается использовать язык программирования Python 3.x, так как для него реализовано большое количество библиотек машинного обучения и анализа данных [15-17].

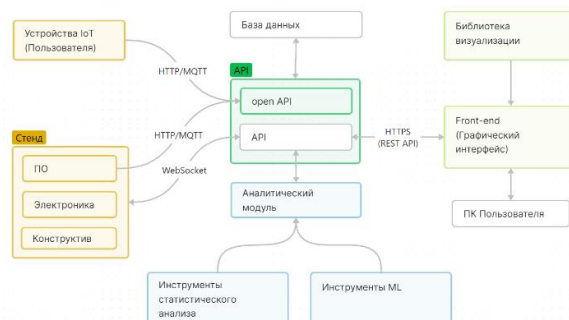


Рис.1. Структурная схема платформы

Описание функционала

Планируется, что платформа IoT Educational DataHub предоставит следующие функции:

1. Сбора данных от различных устройств с поддержкой нескольких протоколов (HTTP, HTTPS, MQTT) и групп устройств.

2. Обработки и анализа полученных данных с использованием встроенных инструментов и алгоритмов.

3. Визуализации данных в виде графиков, диаграмм, таблиц, индикаторов и других форматов.

4. Объединения данных в наборы.

5. Обмена данными между пользователями.

6. Управления актуаторами, входящими в демонстрационно-образовательный стенд.

Функционал демонстрационно-образовательного стенда предполагается следующий:

1. Подключение датчиков к контроллеру модели plug and play [18].

2. Управление устройствами мощностью от 1 до 3 кВт.

3. Сбор данных с датчиков.

4. Отправка данных на платформу IoT Educational DataHub.

5. Возможность добавлять модули.

6. Адаптивная подсветка.

При этом планируется обеспечить допустимость значительного количества циклов сборки стенда. Удаленный доступ может быть организован с использованием очереди и ограничения по времени, как это реализовано в платформах проверки заданий по программированию. Отладочные платы, исполнительные устройства и сенсоры, устанавливаемые на стенд, могут быть различными, в зависимости от потребностей, так, для практикумов IT Академии Samsung может использоваться следующий набор:

1. Импульсный блок питания Robiton TN2000S
2. Отладочная плата STM32 Nucleo F401RE
3. WI-Fi модуль на базе ESP8266(12E)
4. Силовой ключ N-Channel
5. Ёмкостный датчик влажности почвы
6. Четырёхнопочная клавиатура
7. Датчик температуры
8. Датчик влажности
9. Датчик приближения и освещённости
10. Светодиоды

Применимость концепции для практикумов IT Академии Samsung

Трек (учебный курс) Интернет вещей IT Академии Samsung [19] разработан специалистами московского Исследовательского центра Samsung для студентов технических специальностей. Курс предполагает большое количество практических занятий (кейсов), в ходе которых студенты должны ознакомиться с тем, как работают компоненты IoT, а также сделать итоговый проект.

Предлагаемая программно-аппаратная платформа может быть интегрирована в существующий курс и позволит наглядно и удобно работать с данными, поступающими от сенсоров. Помимо использования

во время очных занятий, платформа способна оптимизировать процесс обучения в дистанционном и гибридном формате, что позволит взаимодействовать с устройствами IoT и получать данные от них студентам из разных мест. Это снимает со студентов необходимость очно присутствовать в лаборатории и иметь физический доступ к оборудованию для выполнения любых практических заданий.

Рассмотрим сценарии применения предлагаемой платформы для конкретных заданий IT Академии Samsung.

Суть большей части заданий из кейса 1 сводится к созданию прошивки для платы STM Nucleo с использованием Mbed OS 6 или Mbed OS 5 (которые являются, по сути, операционными системами реального времени [20]). Например, создание прошивки электронного замка. В ходе выполнения кейса студенты сталкиваются с такими организационными проблемами, как ограниченное количество комплектов оборудования и трата существенного времени на подключение периферии к плате. Платформа IoT Educational DataHub позволит существенно оптимизировать этот процесс: для того, чтобы проверить работоспособность прошивки, не нужно будет ожидать своей очереди для получения комплекта оборудования и тратить время на сборку; можно будет просто подключиться к платформе и проверить работу созданной программы с помощью уже подключенных устройств, входящих в образовательный стенд.

В кейсе 2 студентам предлагаются практические задания по сбору данных с сенсоров, подключенных к отладочной плате серии STM Nucleo. В кейс можно было бы добавить расширенную версию задания с обработкой данных, поступающих с датчиков в течение длительного времени. В таком случае речь идет про обработку больших данных; хранить их локально по меньшей мере затруднительно. С помощью платформы IoT Educational DataHub можно собирать данные в течение длительного времени, после чего применить инструменты для анализа больших данных, уже реализованные на платформе.

Дальнейшие кейсы связаны с прямой работой с микроконтроллерами серии ESP32 и ESP8266 или подключением их к отладочной плате STM Nucleo. Проблемы остаются теми же: из-за существенной стоимости оборудования количество комплектов на группу студентов ограничено, что сильно замедляет процесс обучения и выполнения заданий. С помощью платформы вопрос с недостатком оборудования решается: работоспособность прошивки можно будет проверить, не затрачивая время на подключение оборудования и очередь за ним, просто подключившись к платформе.

Итак, IoT Educational DataHub позволит более продуктивно распоряжаться временем практических занятий, а также решит существующие проблемы из-за ограниченного количества комплектов оборудования и постоянной необходимости его физического подключения.

Заключение

В работе рассмотрена возможность создания программно-аппаратной платформы для реализации образовательных программ в сфере Интернета вещей.

Были рассмотрены сценарии применения предлагаемой платформы для выполнения практических кейсов ИТ Академии Samsung. IoT Educational DataHub позволит проверять корректность работы созданных прошивок, не сталкиваясь с такими проблемами, как недостаток комплектов оборудования и необходимость траты времени на физическое подключение устройств. По нашей оценке, платформа применима в каждом кейсе ИТ Академии (за исключением немногих отдельных заданий, не включающих в себя работу с микроконтроллерами, или тех, в которых сама суть задания состоит, например, в правильном подключении периферии) и способна в несколько раз снизить потребность в оборудовании (вплоть до полной замены текущих комплектов оборудования в некоторых кейсах образовательным стендом платформы).

Платформа IoT Educational DataHub упростит процесс получения, обработки и анализа данных, который может быть сложным и трудоемким, в образовательных и исследовательских организациях, в условиях ограниченности ресурсов. Платформа может применяться в условиях очного, дистанционного и гибридного образования.

Список литературы

1. Application of the Internet of Things (IoT) in Health: The Future of Personal Care / С. Н. Sumerli, N. D. Erlinawati, D. Anurogo [и др.]. - Текст : электронный // Journal of World Future Medicine Health and Nursing. - 2024. - № 2. - С. 92-105.
2. Россия онлайн? Догнать нельзя отстать / Бартоломео Банке [и др.] // The Boston Consulting Group (BCG) [Электронный ресурс] – URL: https://image-src.bcg.com/Images/BCG-Russia-Online_tcm27-152058.pdf (дата обращения: 10.11.2020)
3. Number of IoT connected devices worldwide 2019-2023, with forecasts to 2030 / Lionel Sujay Vailshery. // [Электронный ресурс] – URL: <https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/> (дата обращения: 10.02.2024)
4. Методическое пособие «Разработка умных устройств на базе Arduino» / Сомов А.С., Лыжин И.Г. – М: Сколковский институт науки и технологий, 2020. – 80 с.
5. Internet of Things (IoT) Applications in Education: Benefits and Implementation Challenges in Ghanaian Tertiary Institutions / D. K. Dake, G. Bada, A. Dadzie. // Journal of Information Technology Education: Research. – 2022. – № 22. – С. 311-338.
6. Building an IoT Data Hub with Elasticsearch, Logstash and Kibana / Bajaj, Marcin. – Prague : 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), 2017.
7. Yandex IoT Core [Электронный ресурс] – URL: <https://yandex.cloud/ru/services/iot-core> (дата обращения: 25.05.2024)
8. MTC IoT HUB [Электронный ресурс] – URL: <https://developers.mts.ru/mts-iot-hub> (дата обращения: 25.05.2024)
9. Arduino Education [Электронный ресурс] – URL: <https://www.arduino.cc/education> (дата обращения: 25.05.2024)
10. An Introduction to Regression Analysis / Alan O. Sykes. – Coase-Sandor Institute for Law & Economics. Working Paper No. 20, 1993.
11. Random Forests / Breiman, L. – Machine Learning 45, 5–32, 2001. DOI: <https://doi.org/10.1023/A:1010933404324>
12. Decision Tree and Random Forest Implementations for Fast Filtering of Sensor Data / S. Buschjäger, K. Morik. – IEEE Transactions on Circuits and Systems–I: Regular Papers, Vol. 65, No. 1, 2018.
13. Identification of Internet of Things (IoT) Attacks Using Gradient Boosting: A Cross Dataset Approach / Garg, Shruti & Kumar, Vivek & Payyavula, Srinivasa. – ISSN: 1856-4194, 2022.
14. Gradient Boosting for Health IoT Federated Learning / S. Wassan, B. Suhail, R. Mubeen and others. – Sustainability, 2022. DOI: <https://doi.org/10.3390/su142416842>
15. Machine Learning Frameworks for Predictive Analytic / M. Pasupuleti. – 2024.
16. An Overview of Python Libraries for Data Science / A. Joshi, H. Tiwari. – Journal of Engineering Technology and Applied Physics. 5. 85-90, 2023.
17. Review and comparative analysis of machine learning libraries for machine learning / M. Gevorkyan, A. Demidova and others. – Discrete and Continuous Models and Applied Computational Science, 2019. DOI: 10.22363/2658-4670-2019-27-4-305-315.
18. Plug and Play / TechTerms [Электронный ресурс] – URL: https://techterms.com/definition/plug_and_play (дата обращения: 10.02.2024)
19. Трек “Интернет вещей”. ИТ Академия Samsung. [Электронный ресурс] – URL: <https://innovationcampus.ru/internet-of-things> (дата обращения: 10.02.2024)
20. Mbed OS. [Электронный ресурс] – URL: <https://os.mbed.com> (дата обращения: 10.02.2024)

ИССЛЕДОВАНИЕ РЕАЛИЗАЦИИ КРИПТОГРАФИЧЕСКОГО РАСШИРЕНИЯ АРХИТЕКТУРЫ RISC-V

Фролов К.Д., Мирошниченко Л.И.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В данной работе рассматриваются пути к реализации криптографического расширения свободной архитектуры RISC-V и потенциальный прирост производительности при задействовании данного рас-

шифрования в программах, использующих алгоритмы шифрования.

Введение

RISC-V – свободная архитектура, описывающая набор инструкций, курируется международной организацией RISC-V International, не имеет закрытых патентов и не требует лицензионных отчислений для использования. Основной особенностью данной архитектуры является модульная структура, предлагающая минимальный базовый набор инструкций и возможность наращивать функционал при помощи дополнительных расширений [1].

Криптография играет значительную роль в современной цифровой среде, поскольку она обеспечивает защиту данных при их хранении и передаче, особенно в контексте распространенной сетевой коммуникации и других аспектов современной жизни. Криптографические алгоритмы могут быть реализованы не только программно, но и аппаратно, что может дать существенный выигрыш во времени работы алгоритма. Целью данной работы является формирование информационной базы для дальнейшей реализации криптографического расширения архитектуры RISC-V.

В данной работе представлен обзор открытых ядер RISC-V, которые могут быть потенциально интересны для внедрения криптографического расширения, краткий обзор популярных алгоритмов шифрования, а также дана приблизительная оценка возможного ускорения работы алгоритма.

Обзор ядер

Основными критериями для отбора подходящих для данной работы ядер являются открытая лицензия, Verilog и SystemVerilog, как языки описания аппаратуры, полнота реализации базового набора инструкций, а также наличие исчерпывающей документации. По данным критериям были отобраны следующие ядра:

1. CVA6;
2. CV32E40P;
3. DarkRISCV;
4. SCR1.

Процессор CVA6 представляет собой 64-разрядное ядро с 6-ступенчатым конвейером. Он реализует 3 уровня привилегий M, S, U для поддержки полноценных операционных систем, таких как Unix. Отличительной особенностью данной реализации является наличие 6 стадий конвейера, в отличие от классического конвейера с 5 стадиями. Еще одной особенностью является наличие PC Generation, который отвечает за вычисление адреса следующей инструкции и учет ветвлений. Он включает специальные блоки, такие как BHT (Branch History Table) для хранения истории ветвлений PC и BTB (Branch Target Buffer) для буферизации ветвлений. PC Gen также прогнозирует будущие ветвления на основе истории. Кроме того, процессор поддерживает работу с CSR-инструкциями, для чего имеется отдельный буфер, поскольку инструкция CSR изменяет состояние ядра [2].

CV32E40P – это софт-процессорное ядро, разработанное OpenHWGroup на SystemVerilog, обладает 32-битной архитектурой.

Ядро реализует RV32I, а также стандартные расширения C (расширение для сжатых инструкций), M (расширение для целочисленного деления и умножения), Zicntr (счетчики производительности), Zicsr (инструкции для контроля и регистра статуса), Zifencei (механизм получения инструкций) [3]. Дополнительно могут быть подключены расширения F (для операций с вещественными числами одинарной точности) и Zfinx (X регистры для расширения F).

Ядро имеет опциональные нестандартные расширения Xcv (CORE-V PULP ISA) и Xcvelw (CORE-V PULP Cluster ISA) для работы с PULP (Parallel Ultra-Low-Power).

DarkRISCV – софт-процессор, разработанный на Verilog, реализующий RV32I и RV32E. Ядро обладает небольшими размерами, что позволяет с меньшими трудностями разобраться в ядре и модифицировать его. Имеется поддержка синтеза для ПЛИС от Xilinx, а также некоторых плат Intel FPGA [4].

Ключевые возможности ядра:

1. Поддержка компиляции программ с GCC;
2. Поддержка UART-интерфейса;
3. Наличие контроллера прерываний;
4. Наличие контроллера кэшей;
5. Наличие GPIO и таймера;
6. Наличие предсказателя ветвления.

Ядро SCR1 от российской компании Yadro реализует расширения RV32I и RV32E, а также дополнительные расширения RVM (Risc-V Multiplication extension) и RVC (Risc-V Compressed extension) [5]

Основные особенности ядра:

1. Наличие 2-4 ступенчатого конвейера;
2. Наличие дополнительного встроенного программируемого контроллера прерываний с 16 линиями IRQ;
3. Наличие дополнительной подсистемы отладки с интерфейсом JTAG;
4. Наличие 32-разрядного внешнего интерфейса AXI4/AMB-Lite;
5. Наличие 3-х рекомендуемых преднастроенных конфигураций ядра.

Обзор алгоритмов шифрования

AES (Rijndael) [6] – алгоритм симметричного шифрования, был создан в 1998 году, все еще активно применяется в различных сферах информационных технологий. Алгоритм AES основан на замене байтов, сдвигах строк и столбцов, а также применении операций XOR над байтами внутри блока данных. Он использует ключ для шифрования и дешифрования данных; ключ может быть различной длины: 128, 192 или 256 бит. Алгоритм обладает высокой стойкостью к криптоанализу и считается надежным методом шифрования.

SHA-2 [7] – семейство криптографических хэш-функций, используется для генерации фиксированных хэш-значений из входных данных произвольной длины. Эти хэш-значения обычно используются для проверки целостности данных, аутентификации и

обеспечения безопасности информации. Алгоритм обладает высокой стойкостью к криптоанализу и широко применяется в различных приложениях.

Потенциальный прирост производительности

Программные реализации алгоритмов шифрования требуют несколько инструкций для одного цикла шифрования, в то время как аппаратная реализация может реализовывать те же операции в одну инструкцию (рис. 1).

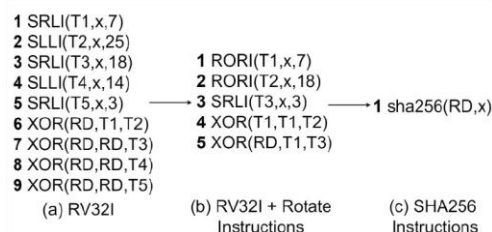


Рис. 1. Число используемых инструкций в зависимости от набора команд [8]

При использовании криптографического расширения может быть достигнуто увеличение скорости выполнения программы в 1,5 и более раз. Также данное расширение обеспечивает уменьшение объема использованной памяти от 1,2 раз [8].

Заключение

В этой работе рассмотрены перспективные софтверные ядра, которые могут быть дополнены криптографическим расширением, криптографические алгоритмы, которые являются объектом расширения, и потенциальные выгоды такого расширения.

Список литературы

1. Specifications // RISC-V International. URL: <https://riscv.org/technical/specifications/> (дата обращения: 07.02.2024).
2. CVA6: An application class RISC-V CPU core // CVA6 User Manual. URL: <https://docs.openhwgroup.org/projects/cva6-user-manual/> (дата обращения: 08.02.2024).
3. OpenHW Group CV32E40P User Manual // CV32E40P User Manual. URL: <https://docs.openhwgroup.org/projects/cv32e40p-user-manual/en/latest/> (дата обращения: 08.02.2024).
4. DarkRISCV Repository // GitHub. URL: <https://github.com/darklife/darkriscv> (дата обращения: 11.02.2024).
5. SCR1 Repository // GitHub. URL: <https://github.com/syntacore/scr1> (дата обращения: 11.02.2024).
6. Daemen J., Rijmen V. AES proposal: Rijndael. – 1999.
7. Dang Q. H. Secure hash standard. – 2015.
8. Nişancı G., Flikkema P. G., Yalçın T. Symmetric Cryptography on RISC-V: Performance Evaluation of Standardized Algorithms // Cryptography. – 2022. – Vol. 6. – No. 3. – Art. No. 41.

ЦИФРОВЫЕ ДВОЙНИКИ НА БАЗЕ РАЗВИТИЯ ТЕХНОЛОГИЙ ВІМ И ІОТ

Тимошкевич И.В.

Белорусский государственный университет информатики и радиоэлектроники, кафедра информационно-коммуникационных технологий

Аннотация

В этой статье проведен анализ взаимодействия совместной работы информационного моделирования зданий и интернета вещей и их возможностей для создания цифрового двойника. Перенос данных с Интернета вещей в Информационную модель позволяет сделать процесс анализа состояния здания более удобным, а также помогает в создании цифрового двойника.

Ключевые слова: ВІМ, ІоТ, строительная индустрия, ЖКХ, эксплуатация.

Введение

Цель работы — возможность формирования цифрового двойника, изучение основных возможностей взаимодействия Информационного моделирования зданий (Building Information Modeling, далее, ВІМ) [1] и Интернета вещей (Internet of Things, далее, ІоТ) [2]. Информационное моделирование зданий и Интернет вещей могут работать вместе, это помогает повысить удобство управления зданиями, а также комфорт и безопасность жителей. Актуальность вызвана развитием взаимодействия данных технологий для получения двойника модели определенного сооружения, в котором применяются счетчики и контроллеры с технологией Интернета вещей.

Цифровой двойник (англ. Digital Twin) — цифровая копия физического объекта или процесса, помогающая оптимизировать эффективность бизнеса. Концепция «цифрового двойника» является частью четвертой промышленной революции и призвана помочь предприятиям быстрее находить физические проблемы, точнее предсказывать их результаты и производить более качественные продукты. В нашем случае интерес прикован к объектам жилищно-коммунального хозяйства, а именно к многоквартирным домам.

Обсуждение взаимодействия интернета вещей с цифровой моделью.

ВІМ — это процесс создания цифрового представления здания, которое включает в себя информацию о его конструкции, материалах, системах отопления, вентиляции, кондиционирования воздуха, освещении и электричестве. Эта информация может быть использована для создания виртуальной модели здания, которая пригодится для планирования, проектирования, строительства и эксплуатации зданий [1]. Принципиальная схема ВІМ представлена на рис. 1.



Рис. 1. Схема принципа работы BIM-технологии

IoT — это сеть физических объектов, которые подключены к интернету и могут обмениваться данными. В контексте зданий IoT может включать в себя датчики, которые могут измерять температуру, влажность, качество воздуха, уровень освещенности и другие параметры окружающей среды в здании [2]. Принципиальная схема IoT представлена на рис.2.

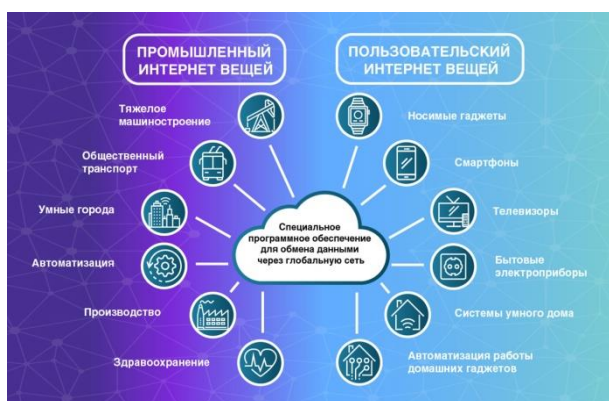


Рис. 2. Схема работы интернета вещей

BIM и IoT могут работать вместе, чтобы улучшить управление зданиями и повысить комфорт и безопасность жителей. Например, используя данные из модели BIM, можно разместить датчики IoT в стратегических местах здания, чтобы собирать информацию о параметрах окружающей среды. Эти сведения могут быть использованы для оптимизации работы систем отопления, вентиляции, кондиционирования воздуха, освещения и электричества, что может привести к снижению затрат на энергию и повышению комфорта жителей. Кроме того, данные IoT будут полезны для мониторинга состояния здания и его систем, что может помочь в предотвращении проблем и сокращении времени простоя [3].

Две технологии объединяются, производя революционный скачок в мире строительства. BIM и IoT — это технологии, которые меняют как строительство зданий, так и взаимодействие с окружающей средой.

Совместная работа, существующая между BIM и IoT, позволяет взаимодействовать со всеми элементами здания и объединять их, дает возможность реагировать на их изменения.

Бесшовное слияние информации и функциональности, объема данных и дизайна, который оживляет

здания, — вот, что происходит, когда BIM и IoT объединяются. Интеграция между двумя технологиями создает мощный синергетический эффект, улучшающий процесс строительства. Основные задачи взаимодействия:

1. Сбор и анализ данных: Устройства Интернета вещей собирают со строительной площадки в режиме реального времени различные данные, включая температуру, влажность и состояние оборудования. Эта информация включается в 3D-модель с помощью BIM, что позволяет проводить немедленный анализ и принимать решения.

2. Автоматизация и контроль: Интерфейс BIM позволяет управлять устройствами Интернета вещей, что помогает автоматизировать различные строительные процессы. Например, изменять системы кондиционирования воздуха или освещения в зависимости от факторов окружающей среды, таких как заполняемость помещения людьми.

3. Сотрудничество и коммуникация: IoT и BIM обеспечивают эффективное взаимодействие между различными заинтересованными сторонами. Обмен информацией в режиме реального времени обеспечивает хорошую диспетчеризацию и даёт гарантию, что информацию получают все.

4. Техническое обслуживание и управление объектами: BIM и IoT продолжают сотрудничать в управлении объектами после завершения строительства. IoT-устройства помогают в предиктивном обслуживании, контролируя и управляя различными системами, а BIM предлагает детальную модель здания.

Суть в том, чтобы получать информацию от счетчиков Интернета вещей, переносить их в информационную модель и в онлайн-режиме в цифровом двойнике здания анализировать состояние теплопотерь, теплопотребления дома.

Этап проектирования:

1. Анализ энергоэффективности: BIM-модели с интегрированными IoT-датчиками могут моделировать энергопотребление, помогая проектировать здания с низким энергопотреблением.

2. Планирование безопасности объекта: IoT-устройства отслеживают состояние участка, а BIM использует эту информацию для планирования безопасных методов строительства.

Этап строительства:

1. Мониторинг в реальном времени: Обратная связь в реальном времени от датчиков IoT используется для обновления в BIM-модели информации о том, как идет строительство.

2. Контроль качества: интеграция BIM и IoT улучшает контроль качества, обеспечивая соблюдение проектных спецификаций.

Этап эксплуатации:

1. Управление объектом: для эффективного управления объектом BIM и IoT предлагают комплексное представление о системах здания.

2. Предиктивное обслуживание: BIM предоставляет информацию, необходимую для технического обслуживания, сокращая время простоя, а IoT выявляют потенциальные проблемы.

Трансформация M2M

Трансформация M2M (Machine-to-machine) — «от машины к машине» — в IoT создает основу для обратной связи по данным для цифровых близнецов (см. рис.3), так как перестает быть точечной и позволяет говорить об онтологической транспарентности и уже реализовывать ее в рамках европейских онтологических стандартов IoT, принятых в 2019 году [4]. Это первые такого рода стандарты, которые пригодны для построения мультидоменных онтологий цифровых двойников.



Рис.3. Трансформация M2M в IoT

В публикации [4] приводится обзор новых европейских онтологических стандартов IoT. Они фактически создают часть условий, в которых появляются цифровые близнецы. Следующим этапом в IoT является не просто установка системы датчиков, а возможность виртуального представления физического мира с помощью цифрового близнеца, который моделирует отношения между людьми, устройствами и самим пространством. Стоит отметить, что 75% организаций, внедряющих технологию IoT, уже используют цифровых близнецов как часть своей системы или планируют сделать это в течение года.

Один из примеров, который использует представитель Azure, — это программирование термостата. Если вы хотите измерить и поддерживать температуру офисного здания, то, например, движение и количество людей внутри него будут играть большую роль в этом. Цифровой близнец может помочь отобразить эти изменения».

На рис.4 представлен пример возможности взаимодействия BIM и IoT систем.



Рис.4. Взаимодействие BIM и интернета вещей

Ценностное предложение: BIM и IoT в действии

Объединение IoT и BIM — это не просто технологический прогресс, это ориентированная на ценности стратегия, которая соответствует стремлению строительной отрасли к качеству, сотрудничеству и

инновациям. Ценностное предложение BIM и IoT заключается в следующем:

1. Эффективность: Автоматизация, интеграция данных в режиме реального времени, оптимизация процессов - все это сокращает время и расходы на строительство.

2. Сотрудничество: Улучшение коммуникации между заинтересованными сторонами способствует сотрудничеству, обеспечивая единство взглядов.

3. Инновации: Инновационные возможности в проектировании, строительстве и обслуживании становятся возможными благодаря интеграции BIM и IoT.

4. Обеспечение качества: Приверженность качеству демонстрируется мониторингом и контролем в режиме реального времени, которые обеспечивают соблюдение стандартов.

Практический пример

Физический близнец отражается в цифровом двойнике с помощью развивающейся технологии IoT, а также в развитии связи, а именно 5G. Такие факторы, как климат и движение, могут отслеживаться датчиками и отображаться в экосистеме цифрового близнеца, чтобы создать максимально точную и реалистичную копию.

Хорошо, когда цифровые близнецы превосходны как технология, это способность учитывать контекст существующих структур алгоритмов и данных. Как только близнец будет точно создан, можно проводить цифровое моделирование, которое может предотвратить проблемы, возникающие в физическом близнеце, в то время как изменения могут быть опробованы с помощью цифрового близнеца и позволят возвращать «реалистичные» результаты.

usBIM.IoT — это масштабируемая и настраиваемая система управления BIM, способная отслеживать операции, связанные с инфраструктурой. Например, с помощью датчиков Интернета вещей можно управлять событиями или физическими изменениями внутри зданий, мостов и автомагистралей. Возможность доступа к информации из любого места и в любое время также приносит большие преимущества, такие как экономия средств на техническое обслуживание, снижение затрат на электроэнергию или улучшение качества жизни в городе благодаря интеллектуальным уличным фонарям и контролю за дорожным движением.

BIM (информационная модель здания) позволяет создавать цифровую модель, которая представляет цифрового двойника выполненных или подлежащих выполнению работ.

Реальной моделью или строительным объектом можно управлять через его цифрового двойника (модель объекта в облачном сервисе) с помощью чрезвычайно передовых методов, которые также используют интеграцию с датчиками интернета вещей.

Датчики технологии IoT представляют собой довольно простые устройства, но, применяемые к зданиям и инфраструктуре, могут генерировать очень высокий инновационный потенциал с точки зрения управления как технологическими, так и финансо-

выми аспектами для их правильной и эффективной работы.

Подключенные к специальным устройствам, они могут измерять температуру, уровень влажности, интенсивность освещения, движения или любые другие типы механической или физической информации. Объединяя эти результаты с BIM-технологиями, результирующая обратная связь представляет связь в режиме реального времени между обнаруженными данными и 3D-моделью (также используемой в визуальных целях), открывающую бесконечное количество возможностей с точки зрения эффективности и управления проектами.

При использовании данных с IoT систем есть возможность постоянного онлайн обновления документов процесса строительства или эксплуатации. А также с помощью BIM-модели, ведется управление в удаленном режиме, даже в критические моменты, и есть возможность оптимизировать весь процесс строительства от полученных нами данных в реальном времени удаленно, с помощью интернета.

Фактически, эти датчики интернета вещей поддерживают постоянную связь между физической средой и цифровой моделью, открывая новые возможности в области управления проектами. Наглядный пример взаимодействия Интернета вещей с информационным моделированием представлен на рис.5.



Рис.5. Цифровой двойник с IoT-информацией

Этот огромный массив данных представляет собой вспомогательную основу для каждого этапа принятия решений и управления строительным процессом: от строительной площадки до управления и технического обслуживания с течением времени.

Например, на этапе строительства датчики позволяют вам получать все необходимые данные, касающиеся присутствия персонала на строительной площадке, рабочих потоков, пикового присутствия и любых перекрывающихся задач. Они могут собирать в режиме реального времени информацию об использовании машин, оборудования и материалов, получать постоянно обновляемый статус хода работ и расходных материалов.

Благодаря облачным платформам эта огромная база данных естественным образом доступна всем, кто участвует в проекте, в соответствии с идеально согласованным подходом с философией информационного моделирования зданий.

Для реализации знаний в области интернета вещей в BIM можно воспользоваться онлайн-платформой usBIM.IoT, интеграцией BIM и интернета вещей, которая использует информацию BIM-модели для создания автоматической связи между физическим устройством интернета вещей и его цифровым двойником в рамках рабочих процессов BIM.

Все данные, записанные датчиками, отправляются в режиме реального времени в приложение usBIM, которое получает, анализирует и отображает информацию, соответствующим образом преобразованную в BIM-модели в открытом формате IFC.

BIM-модель оживает за счет изменения своих характеристик на основе данных, передаваемых каждым датчиком, а визуальные изменения цифрового двойника в BIM в режиме реального времени (на основе значений физического устройства) позволяют осуществлять точный контроль даже в больших средах с десятками тысяч устройств.

Благодаря трехмерной модели и объединенной системе визуализации, предлагаемой BIM, вы получаете разбивку экосистемы Интернета вещей на подсистемы и навигацию сверху вниз.

Заключение

Цифровые близнецы открывают совершенно новый способ поиска оптимизации в физических пространствах. Умные здания и инфраструктуры рассматриваются как идеальный образец технологии, а снижение затрат на инфраструктуру через развитие 5G, IoT, BIM облегчает их разработку в сторону цифровых двойников.

Цифровой двойник был устоявшейся концепцией на рубеже 20 и 21 веков, хотя ограничения сдерживали его как достойную инвестицию. Однако в последнее время развитие Интернета вещей сделало цифрового близнеца жизнеспособным настолько, что он оказывает влияние на широкий круг различных отраслей. В случае умного здания сооружение не может быть умным без цифрового близнеца, который является программным обеспечением для управления зданием. Следовательно, здание и цифровой близнец — это почти одно и то же: близнецы не могут быть разделены и не могут существовать самостоятельно. В конечном счете нет никакого способа создать действительно совершенного цифрового близнеца, по крайней мере, пока. IoT-датчики точны, но существует множество факторов, влияющих на любой продукт в любой момент времени, чтобы безупречно отобразить его без труда.

Таким образом, формирование цифрового двойника, изучение основных возможностей взаимодействия BIM и IoT, их совместная работа улучшают возможность управления зданиями и повышают комфорт и безопасность жителей.

Способность полностью реализовать потенциал цифровых близнецов в различных приложениях и отраслях зависит от доступа и достоверности данных, поступающих в них. Предполагалось, что сведения, используемые для взаимодействия между объектами, могут привлечь злоумышленников, кото-

рые хотят либо заразить, либо нарушить обмен. Интеграция блокчейна может предложить решение. Высокая пропускная способность и низкая задержка от 5G улучшат сбор данных и доступ к ним в процессе доставки проекта. Повышенная наглядность данных обеспечивает принятие решений на этапе проектирования, помогает минимизировать проблемы и изменения во время строительства и потенциально уменьшает будущие ремонтные работы.

Однако стоит вспомнить историю интеграции BIM в цифровую экономику, нынешняя трансформация BIM в цифрового двойника потребует осознанного и рассчитанного риска при комплексном внедрении в крупных инфраструктурных проектах, но множество выгод (и не только экономических) того стоят.

Список литературы

1. BIM – краткий обзор и основные понятия. // Информационный BIM портал. [Электронный ресурс] — <https://bim-portal.ru/stati/kratkij-obzor-bim/> Дата обращения 10.01.2024.
2. Суомалайнен А. Интернет вещей: видео, аудио, коммут. // ДМК Пресс, 2019. – 120с.
3. Building Information Modeling (BIM) And The Internet of Things (IoT) - How They Work Together // An online resource. [Electronic resource] — <https://www.ddg.global/post/building-information-modeling-bim-and-the-internet-of-things-iot-how-they-work-together> Дата обращения 15.01.2024.
4. Kupriyanovsky, Vasily, et al. "On the new IoT generation-ETSI ontology standards and specifications." International Journal of Open Information Technologies 7.9 (2019): 73-81.

ИССЛЕДОВАНИЕ АРЕНДНЫХ РЕШЕНИЙ НА РОССИЙСКОМ РЫНКЕ ТЕЛЕКОММУНИКАЦИЙ

*Буслаев Д.П., Ельцов М.Е., Портной С.Л.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе проведен обзор современных арендных решений на российском рынке телекоммуникаций, таких как частные сети LTE/5G, мобильные виртуальные операторы и совместное использование радио инфраструктуры. Рассматриваются особенности каждого из приведенных решений, их влияние на инфраструктуру связи, а также перспективы развития в контексте текущих и будущих требований отрасли.

Введение

В свете быстро меняющейся технологической среды и увеличивающейся конкуренции в секторе телекоммуникаций, особое внимание уделяется инновационным подходам к аренде инфраструктуры [1,

2], обеспечивающим оптимальное соответствие потребностям операторов связи и конечных пользователей. В работах [1, 2, 3] частные сети LTE/5G и совместное использование RAN описывается как одна из наиболее перспективных отраслей телекоммуникаций. Данная работа позволяет получить представление о текущем состоянии и перспективах развития рассматриваемых арендных решений в российской телекоммуникационной индустрии. Анализируются преимущества и недостатки различных моделей аренды, а также отдельные аспекты использования их в РФ. Целью работы является обзор существующих арендных решений и обоснование целесообразности использования описываемых арендных решений в рамках развертывания сетей новых поколений. В рамках исследования используется метод анализа существующих арендных решений с уклоном на использование в современных сценариях развития телекома в Российской Федерации.

Частные сети рLTE/5G

Построение частных сетей LTE/5G на объектах критически важной инфраструктуры является одним из самых эффективных решений, предпринимаемых предприятиями для увеличения степени автоматизации производства, а также обеспечения высокого уровня информационной безопасности.

Внедрение подобных технологий позволяет организовать закрытый контур на территории объекта с возможностью использования дополнительных сервисов для повышения цифровизации производства и обеспечения всесторонней безопасности [1].

Архитектура сети рLTE представлена на рис.1. Для обеспечения услуг МСРТТ (Missional Critical Push-to-Talk) на стороне приложений расположен соответствующий сервер. Под UE (User equipment) в данном случае подразумеваются не только обычные смартфоны сотрудников, но и оборудование для производства, датчики IoT, рации МСРТТ и др.

Все сетевое оборудование, начиная от ядра сети, заканчивается подсистемой базовых станций закупается и устанавливается на объекте самим предприятием в сотрудничестве с вендором. Однако в связи со спецификой выдачи лицензированных частот в Российской Федерации [4], организации необходимо вступать в сотрудничество и арендовать частоты у крупных телекоммуникационных операторов, которые будут заниматься вопросами законодательного регулирования с государственными органами. Это сотрудничество расширяется в связи с тем, что часть оборудования управления RAN (Radio access network) должна находится на стороне оператора, поскольку владельцы частот обязаны контролировать выведенные в эфир базовые станции (БС).

Такой вариант взаимодействия является недостатком данных решений и часто может не устраивать владельцев предприятий, которые хотят организовать единый закрытый контур с полной независимостью от оператора связи.

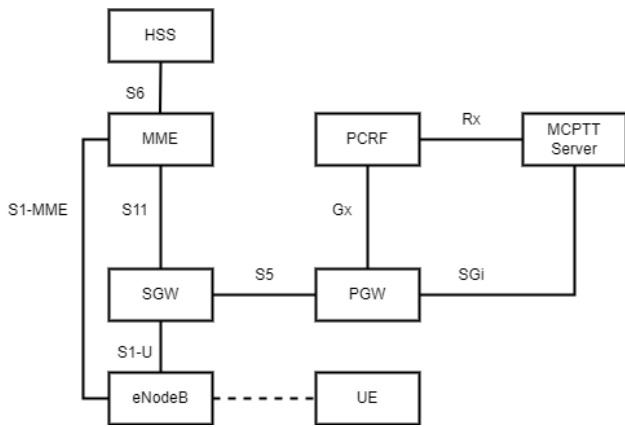


Рис.1. Архитектура частной сети LTE

Примером эффективной работы частых сетей LTE/5G являются объекты горнодобывающей промышленности, такие как шахты и карьеры. Эти объекты обычно охватывают большие территории с неровным рельефом, где действуют мобильные транспортные средства и рабочие, а также развернуты различные промышленные системы, требующие сетевого взаимодействия.

Нахождение на предприятии сравнительно небольшого количества БС позволяет легко проводить модернизацию сети до технологий 5G – достаточно просто установить новые модели БС, поддерживающие 5G и активировать соответствующий функционал. Распределение технологий, используемых на предприятиях для обеспечения связи, представлено на рис.2.

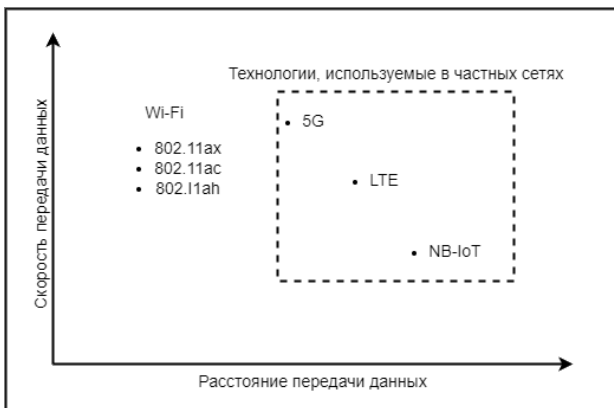


Рис.2. Технологии, используемые в частных сетях

К концу 2023 года в России реализовано 20 коммерческих сетей Private LTE/5G-Ready, а также 2 сети в Казахстане [5]. Из них четыре сети принадлежат ОАО "РЖД" и созданы без участия сотового оператора. Эти сети расположены на сортировочных и ж/д станциях Новосибирска и Челябинска. Кроме того, в РФ построено около 40 пилотных проектов с внедрением технологий 5G в частные сети.

RAN sharing

Концепция RAN sharing в сфере телекоммуникаций [2] представляет собой стратегию, при которой несколько операторов мобильной связи совместно

используют одну и ту же инфраструктуру радиодоступа. Это означает, что они могут делиться базовыми станциями, антеннами и другими элементами сети для обеспечения своих услуг связи.

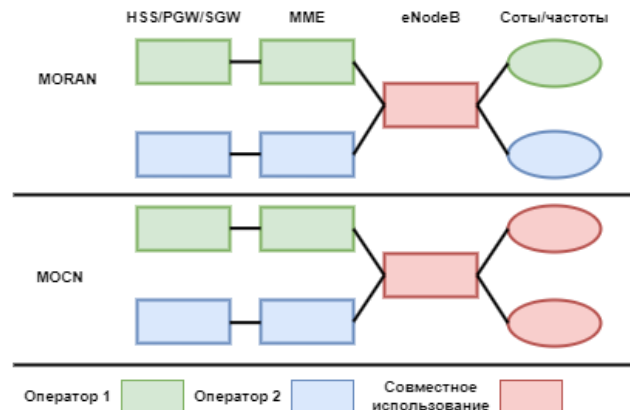


Рис.3. Методы шеринга MORAN и MOCN

Спецификацией консорциума 3GPP [3] определяется несколько моделей совместного использования, из которых хорошо выделяются две из них (см. рис.3):

- MORAN. Совместно используется только RAN. Операторы используют свои частоты.

- MOCN. Вдобавок к RAN, участвующие операторы объединяют свои блоки частот в единую полосу.

Проведем расчет средней пропускной способности базовой станции в моделях MORAN и MOCN [6]:

$$R = S \times W,$$

где S – средняя спектральная эффективность соты (бит/с/Гц), W – ширина канала (МГц).

Средняя спектральная эффективность соты в беспроводных сетях представляет собой меру спектральной эффективности использования радиочастотного спектра в конкретной соте. Данный показатель может меняться в зависимости от используемой схемы MIMO или ширины канала. В наших расчетах будем использовать средние значения для сетей LTE с использованием полного частотного дуплекса [1] - 1,829 бит/с/Гц для канала Uplink и 4,48 бит/с/Гц для канала Downlink.

Тогда средняя пропускная способность БС может быть вычислена как:

$$R_{eNodeB} = R \times N,$$

где N – число секторов БС (в большинстве случаев равно 3). Ширину канала примем за 10 МГц для MORAN и общую при совместном использовании MOCN – 20 МГц.

Для MORAN:

$$R_{eNodeB}(UL) = 1,829 \times 10 \times 3 = 54,87 \text{ Мбит/с}$$

$$R_{eNodeB}(DL) = 4,48 \times 10 \times 3 = 134,4 \text{ Мбит/с}$$

Для MOCN:

$$R_{eNodeB}(UL) = 1,829 \times 20 \times 3 = 109,74 \text{ Мбит/с}$$

$$R_{eNodeB}(DL) = 4,48 \times 20 \times 3 = 268,8 \text{ Мбит/с}$$

Из полученных результатов можно сделать вывод, что вне зависимости от используемой в БС схемы MIMO, средняя пропускная способность по обо-

им каналам для MOCN будет превышать значения для MORAN в 2 раза.

При дальнейшем увеличении охвата зоны совместного использования возможно реализовать модель CWCN (Gateway Core Network), при которой узел управления мобильностью абонентов может использоваться совместно несколькими участвующими операторами, тем самым позволяя значительно сократить сигнализационные сообщения в сети и снизить OPEX и CAPEX. Узел PGW в данном случае используется раздельно, SGW – при необходимости может использоваться в режиме совместного использования.

Модель реализации совместного использования транспортной сети также имеет свои особенности (см. рис.4). На стороне оператора А расположен специальный маршрутизатор агрегации MBAR, принимающий трафик общий трафик от БС. Далее он передается на пиринговые маршрутизаторы, имеющиеся на стороне каждого участвующего оператора. Трафик каждого оператора направляется через виртуальные локальные сети VLAN между MBAR и общей БС, что обеспечивает возможность использовать одну выделенную линию с несколькими VLAN. Маршрутизация данных между сетями участвующих операторов происходит с использованием протокола BGP (Border Gateway Protocol).

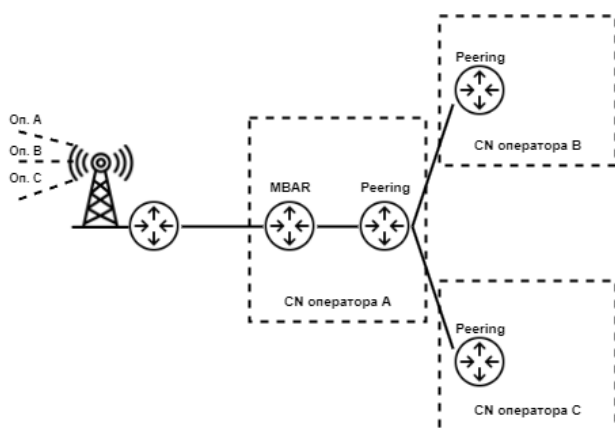


Рис.4. Топология транспортной сети при шеринге

Мобильные виртуальные операторы MVNO

Мобильные виртуальные операторы (MVNO) представляют собой особый сегмент в телекоммуникационной индустрии, функционирующий без собственной инфраструктуры сети. Вместо этого они заключают соглашения с основными опорными операторами мобильной связи MNO для аренды частотных ресурсов и доступа к сети с последующим предложением услуг под собственным брендом.

В 2021 году рынок MVNO в России показал рекордные темпы роста по абонентской базе и выручке. Россия стала лидером мирового роста абонентской базы MVNO, составившим 11% [7]. По данным «ТМТ Консалтинг», в 2020 году в России насчитывалось 12,0 млн абонентов MVNO [8]. Согласно статистике и прогнозам на сайте компании «Теле2», в 2020 году в России было 11,2 млн абонентов виртуальных операторов, и к 2025 году их число может увеличиться до 25 млн [9].

В 2022 году выручка банковских виртуальных операторов, текущих лидеров российского рынка MVNO, выросла на 38% [10] и составили более 8,7 млрд руб. (по сравнению с 6,3 млрд руб. в 2021 году).

Различают несколько типов виртуальных операторов [11], которые классифицируются по размеру зоны ответственности как MVNO, так и MNO. Организация, которая открывает MVNO, может организовать бизнес по агентской схеме, перепродавая закупленные у MNO по оптовым ценам пакеты трафика, или создать собственное ядро сети и сопутствующие сервисы, арендуя только подсистему базовых станций.

Типовая модель реализации Full MVNO включает в себя CN (Core Network) и дополнительные узлы, реализующие вспомогательные услуги и предоставляющие высокий уровень информационной безопасности в сети (см. рис.5). Например, площадка CAMEL gateway, первоначальной задачей которой была организация роуминга, позволяет абонентам MVNO использовать специальные интеллектуальные услуги. Также предусматривается плавное масштабирование сети, которое подразумевает организацию новых сервисов или аппаратное/программное резервирование ядра сети.

Типовая модель реализации Full MVNO включает в себя CN (Core Network) и дополнительные узлы, реализующие вспомогательные услуги и предоставляющие высокий уровень информационной безопасности в сети (см. рис.5). Например, площадка CAMEL gateway, первоначальной задачей которой была организация роуминга, позволяет абонентам MVNO использовать специальные интеллектуальные услуги. Также предусматривается плавное масштабирование сети, которое подразумевает организацию новых сервисов или аппаратное/программное резервирование ядра сети.

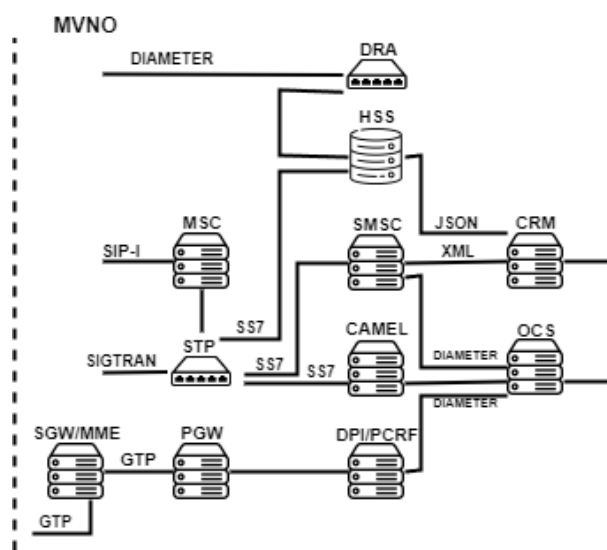


Рис.5. Типовая архитектура Full MVNO

Однако, возвращаясь к закону о связи Российской Федерации [4], стоит отметить, что названия вышеописанных моделей не регламентируются государственными документами. Существует только единая лицензия на предоставление услуг связи как MVNO, особенности которой прописаны в ряде нормативных документов [12]. На данный момент опорные операторы помогают реализовать различные модели MVNO, основываясь на требованиях заказчика, который указывает по какой схеме аренды инфраструктуры он собирается работать.

Заключение

Арендные решения в российской телекоммуникационной индустрии представляют собой динамично развивающийся сектор, играющий ключевую роль в модернизации существующей инфраструктуры. Внедрение новых технологий 5G требует глобальной модернизации текущей инфраструктуры, и именно здесь арендные решения могут стать эффективным инструментом для развертывания сетей нового поколения.

Так, возможность совместного использования RAN может значительно сократить расходы на закупку нового оборудования на ранних стадиях проектов. Ввиду высокой стоимости и нехватки оборудования крупные телекоммуникационные операторы могут заключать соглашения на совместное использование RAN до того момента, пока не появится оптимальное и доступное решение по развертыванию сетей пятого поколения.

Работа по запуску новых проектов частных сетей также сможет внести свой вклад в данном сценарии. Телекоммуникационные операторы уже сейчас запускают множество пилотных проектов [6] частных сетей 5G. Каждый подобный проект является индивидуальным техническим решением, а это означает, что операторы смогут получить всесторонний опыт развертывания сетей пятого поколения и использовать его в своих глобальных проектах.

Список литературы

1. М.И.Бухтеев, С.Л.Портной Архитектура частных сетей lte // Первая миля. - 2023. - №Выпуск #1/2023
2. Е.Зайчик Совместное использование сетевой инфраструктуры операторами связи // ПЕРВАЯ МИЛЯ. - 2019. - №Выпуск #7/2019. - С. 54-59.
3. 3GPP. Network Sharing; Architecture and functional description (Release 15)), TS 23.251 V15.1.0. Technical Specification Group Services and System Aspects, 2018.
4. Закон Российской Федерации "Федеральный закон "О связи" " от 07.07.2003 № N 126-ФЗ // Собрание законодательства Российской Федерации
5. Расчетная часть // studfile URL: <https://studfile.net/preview/2948578/> (дата обращения: 12.02.2024).
6. Корпоративные сети Private LTE/5G в России и Казахстане 2023 // COMNEWS URL: <https://www.comnews.ru/content/229764/2023-11-15/2023-w46/1180/korporativnye-seti-private-lte5g-gossii-i-kazakhstan-2023> (дата обращения: 12.02.2024).
7. VII ежегодная Национальная конференция «MVNO Russia 2022. Виртуальные операторы подвижной радиотелефонной связи в России» // TMT Conference URL: <http://surl.li/txgmk> (дата обращения: 12.02.2024).
8. Российский рынок MVNO: итоги 2020 года // TMT РЕЙТИНГ URL: <http://tmt-consulting.ru/wp-content/uploads/2021/04/%D0%A2%D0%9C%D0%A2-MVNO-2020.pdf> (дата обращения: 12.02.2024).

9. Фабрика MVNO Tele2 // TELE2 URL: <http://surl.li/txgnu> (дата обращения: 12.02.2024).

10. Сектор банковских MVNO показал рост // COMNEWS URL: <https://www.comnews.ru/content/225949/2023-05-04/2023-w18/sekto-bankovskikh-mvno-pokazal-rost> (дата обращения: 12.02.2024).

11. Recommendations document on national roaming access terms and conditions, as well as MVNO access terms and conditions // comcom URL: <https://www.comcom.ge/uploads/other/4/4485.pdf> (дата обращения: 12.02.2024).

12. Закон Российской Федерации "Приказ Минкомсвязи РФ "Об утверждении Требований к оказанию услуг подвижной радиосвязи и радиотелефонной связи при использовании бизнес-моделей виртуальных сетей подвижной радиосвязи и радиотелефонной связи". " от 20.10.2017 № N 570 // Собрание законодательства Российской Федерации

ИНТЕГРИРОВАНИЕ 3D МОДЕЛИ РАСПРОСТРАНЕНИЯ СИГНАЛОВ В СРЕДУ МОДЕЛИРОВАНИЯ ДЛЯ БЕСПИЛОТНОГО И ПОДКЛЮЧЕННОГО ТРАНСПОРТА

*Андрончев А.Д.¹, Насыров Д.Д.²,
Рахимуллина В.Е.², Мкртчян Г.А.², Бушуев Н.И.²*

Национальный исследовательский университет

«Высшая школа экономики»,

¹департамент компьютерной инженерии,

²департамент электронной инженерии

МИЭМ НИУ ВШЭ

Аннотация

Технологии беспилотного и подключенного транспорта предлагаются в качестве решений проблем неэффективности транспортных систем. Однако, прежде чем данные технологии можно будет использовать на дорогах общего пользования, они должны пройти обширные испытания, в том числе с помощью высокоточного моделирования. В настоящее время не существует инструментов, позволяющих проводить совместное моделирование подключенного и беспилотного транспорта, которые позволяли бы учитывать как восприятие сенсорами автомобилей 3D окружающей среды, так и распространение радиосигналов в среде с 3D препятствиями. В данной работе изучены существующие решения, описан инструмент для моделирования окружающей среды и предложен метод для увеличения его точности моделирования распространения сигналов между подключенным транспортом с помощью интеграции модели распространения в 3D окружении.

Введение

Рост населения [1] приводит к увеличению спроса на транспорт, а расширение транспортных систем приводит к увеличению количества дорожно-транспортных происшествий, выбросов парниковых

газов, экономическим и временным потерям из-за дорожных заторов.

Решения этих проблем разрабатываются в области беспилотных и подключенных транспортных средств (AV и CV, соответственно), которые могут дополнять друг друга и, следовательно, быть объединены в подключенные беспилотные транспортные средства (CAV). Достаточное количество данных, получаемых CAV со своих сенсоров и по каналам связи с другими такими транспортными средствами (ТС) позволяет им более точно планировать и контролировать свое движение, и как следствие, это может помочь решить выделенные выше проблемы, исключив человеческий фактор из системы. CAV повышают безопасность дорожного движения [2], [3], а также снижают воздействие транспорта на окружающую среду примерно на 60% [4].

Для тестирования различных аспектов CAV была выявлена потребность в интегрированных средах моделирования (ICM/ISE) [5], и создана архитектура CAVISE [6] (Connected and Automated Vehicle Integrated Simulation Environment) – проект, который объединяет решения с открытым исходным кодом для моделирования дорожного движения и обучения автономных транспортных средств. Архитектура CAVISE использует фреймворк OpenCDA [7] в качестве основы для разработки и использует фреймворк Artery [8] для моделирования автомобильных сетей связи и модель GEMV2 [9] для моделирования распространения сигналов в пространстве. Однако, GEMV2 работает только с двумерным пространством и не позволяет моделировать сценарии с перепадами высот или другими трехмерными свойствами (например, с наличием туннелей, мостов и др.).

В этой статье предлагается оптимизировать архитектуру CAVISE путем интеграции трёхмерной модели Orpal – инструмента моделирования распространения электромагнитного излучения [10], [11], в качестве альтернативы существующей модели GEMV2. В результате будет представлена дополненная архитектура CAVISE, которая обеспечивает взаимозаменяемость между двумя инструментами. Для этого сначала проводится сравнение моделей GEMV2 и Orpal, а затем описывается процесс расширения архитектуры интегрированной среды моделирования и представляются выводы о проделанной работе и направления дальнейших исследований.

Сравнение моделей GEMV2 и Orpal

GEMV2 [9] и Orpal [10], [11] представляют собой две важные технологии, применяемые в области подключенных транспортных средств. В этом разделе рассматриваются фундаментальные принципы работы каждой модели, сравниваются их функциональные возможности и определяется, в каких ситуациях одна модель может быть предпочтительнее другой.

GEMV2 – это модель распространения связи между транспортными средствами (V2V) и между транспортными средствами и инфраструктурой (V2I) [9]. Модель вычисляет детерминированные вариации крупномасштабных сигналов и стохастические вари-

ации мелкомасштабных сигналов в зависимости от количества и размера окружающих объектов. Она использует очертания транспортных средств, зданий и листвы (основных источников, создающих помехи сигналам) для определения типов каналов: прямой видимости (LOS), вне прямой видимости (NLOSv) из-за транспортных средств, вне прямой видимости (NLOSb) из-за плотных объектов (зданий и рельефа местности), вне прямой видимости из-за листвы (NLOSf).

GEMV2 может использовать данные о местоположении транспортных средств из моделей транспортной мобильности (например, SUMO [12]) или треков реального мира (например, собранные с помощью GPS [13]), а также контуры и расположение зданий и растительности, которые находятся в свободном доступе в таких проектах, как OpenStreetMap [14].

Детерминированные модели на основе геометрии [15], такие как GEMV2, являются лучшим выбором, если доступна географическая информация; упрощенные модели, основанные на геометрии, при достижении разумной точности, обеспечивают большой выигрыш в масштабируемости. GEMV2 является хорошим примером модели канала, основанной на геометрии, которая может имитировать распространение сигнала в различных средах (например, на шоссе, в сельской местности, в городах и на сложных перекрестках). Она также позволяет проводить вычисления достаточно быстро, чтобы моделировать автомобильные сети по всему городу с тысячами транспортных средств, взаимодействующих по каналам V2V и V2I.

Orpal – это инструмент для моделирования распространения электромагнитного излучения с трассировкой лучей [16] с использованием графических процессоров, реализованный на фреймворке Veneris [11]. Основным преимуществом Orpal является его интеграция с игровым движком Unity3D, который позволяет быстро и интуитивно создавать пользовательские 3D-среды [10], [11]. Orpal проводит трассировку лучей на графическом процессоре с помощью NVIDIA OptiX, обеспечивая более эффективное моделирование при одновременном использовании центрального процессора для имитации трафика [11].

Модели трассировки лучей, как правило, требуют больших вычислительных затрат, поскольку они моделируют распространение множества лучей в окружающей среде, что требует значительных вычислительных ресурсов, особенно для сложных или больших сценариев. GEMV2 разработан таким образом, чтобы обеспечить баланс между точностью и вычислительной эффективностью. По-прежнему моделируя основные характеристики радиоканала между транспортными средствами, он упрощает вычисления за счет использования геометрических соотношений, что в определенных сценариях может ускорить их по сравнению с подходами детальной трассировки лучей, такими как Orpal. Такая масштабируемость менее приоритетна в средах высокоточного моделирования, таких как CAVISE.

Таким образом, Orpal обеспечивает возможность высокодетализированного моделирования, но с более высокими вычислительными затратами, что делает его подходящим для задач, где детализация имеет решающее значение. GEMV2, с другой стороны, предоставляет эффективное и масштабируемое решение для моделирования коммуникации V2V и V2I. Недостаток того, что GEMV2 способен работать только в 2D-пространстве и не может представлять рельеф и 3D-структуры, будет компенсирован с помощью Orpal, который хорошо подходит для моделирования в 3D-пространстве.

Интеграция Orpal в CAVISE

CAVISE – это ИСМ, состоящая из множества модулей, поэтому существуют ограничения на интеграцию Orpal в существующую архитектуру. В этом разделе оценивается осуществимость интеграции, рассматриваются основные проблемы обсуждаются различные методы интеграции и преимущества и недостатки каждого из них.

Предлагаемая модификация CAVISE предполагает, что Orpal будет интегрирован в существующую архитектуру через Veneris. Таким образом, основной задачей внедрения Orpal в новой версии CAVISE становится интеграция с Veneris. Важно отметить, что модуль OMNeT++ должен быть доступен как для фреймворков Veneris [11], так и для Artery [8], интегрированных в CAVISE.

CAVISE использует Unreal Engine 4 (UE4) в качестве графического симулятора; поэтому сложность интеграции Orpal в CAVISE заключается в переносе Veneris с Unity 3D на UE4. Один из способов решения данной задачи перенести его с помощью сторонних библиотек, таких как Project Exodus [17] или Utu Plugin [18]. Эти плагины позволяют провести этот процесс с меньшими затратами времени и труда, чем ручное переписывание кода; однако результаты получаются менее точными по сравнению с первым вариантом. NVIDIA Omniverse [19] еще один инструмент, который можно использовать для интеграции Orpal с CAVISE.

Далее мы рассмотрим преимущества и недостатки каждого из ранее рассмотренных методов. Переписывание кода – самый надежный метод, однако, если объем кода слишком велик, то задача начинает требовать значительные временные и трудовые ресурсы.

Чтобы перенести работу из Unity в Unreal Engine с помощью Utu Plugin пользователю нужно только выбрать необходимые сценарии. Плагин преобразует их и включает все их зависимости. Его пользовательские интерфейсы также просты и удобны. Кроме того, файл журнала автоматически сохраняется во время каждой операции экспорта и импорта, что облегчает отладку в случае возникновения ошибок.

Инструмент Project Exodus имеет возможность экспортировать как отдельные объекты, так и целые проекты. Однако одним из существенных ограничений этого подхода является то, что он поддерживает только Unity версии 2018 года, а часть плагина Unreal Engine была протестирована для трех версий

UE4, в частности 4.25, 4.24 и 4.23, в то время как симулятор CARLA, на котором основан CAVISE, использует версию 4.26 [20].

NVIDIA Omniverse [19] – это платформа для разработки и совместной работы над 3D-приложениями, основанная на Universal Scene Description (USD) и Material Definition Language (MDL) [21]. Она обеспечивает плавную интеграцию нескольких инструментов 3D-моделирования, таких как Unity 3D и Unreal Engine. Он также поддерживает CARLA, симулятор с открытым исходным кодом для исследований автономного вождения. Однако Omniverse совместим только с CARLA в Windows, в то время как CAVISE разработан для Linux. Таким образом, использование Omniverse с CAVISE потребовало бы написания кода для совместимости с Linux.

Таким образом, необходимо оценить эффективность методов по метрикам средней абсолютной ошибки (MAE) и среднеквадратической ошибки (RMSE) при моделировании траекторий движения автомобилей в CARLA. По этим метрикам будет проводиться сравнение истинных координат траектории транспорта и экспортированных координат на тестовом сценарии Orpal. Допустимый интервал составляет $MAE < 0,05$ и $RMSE < 0,07$. Эксперименты проводились с использованием траекторий движения автомобиля, моделируемых в CARLA. Сравнивались требуемые траектории (из тестового сценария) с полученными траекториями для различных методов переноса данных:

- Ручной перенос: минимальные значения MAE и RMSE.

- Project Exodus: MAE увеличилась на 15%, RMSE увеличилась на 20%.

- Utu Plugin: MAE увеличилась на 10%, RMSE увеличилась на 12%.

- NVIDIA Omniverse: MAE увеличилась на 8%, RMSE увеличилась на 10%.

Эти результаты демонстрируют, что автоматизированные методы переноса ухудшают точность моделирования траекторий автомобиля по сравнению с ручным переносом данных, что является критически важной частью для CAVISE. Так, для достижения высокой точности моделирования потребуются использовать ручной перенос, несмотря на его трудоемкость.

Расширенная архитектура интегрированной среды моделирования

В этом разделе будет более подробно описана система CAVISE и рассмотрены две версии ее архитектуры: одна, использующая только GEMV2, и другая, включающая как GEMV2, так и Orpal. Приведены подробные описания модулей и их взаимосвязей.

В базовой архитектуре CAVISE Artery выступает в качестве платформы для моделирования подключенных ТС, использующих коммуникации V2X. Модуль Middleware используется для сбора и применения данных из OpenCDA использования Artery для моделирования обмена сигналами в системе совместной координации дорожного движения CAV. Artery использует INET [22] и Vanetza [8] для реали-

зации беспроводной связи. INET моделирует протокол беспроводного канального уровня для V2X, а Vanetza предоставляет стек протоколов ETSI ITS-G5. Эти модули позволяют вычислять дальность распространения сигналов, задержку и загруженность. Протокол TraCI, входящий в состав SUMO, позволяет обмениваться данными о местоположении транспортного средства, скорости, угле поворота и данными датчиков между Artery (созданном на базе симулятора OMNeT++) и CARLA для обеспечения синхронного моделирования. Данные с транс-

портных средств, полученные с помощью CARLA, передаются по протоколу TraCI от SUMO в Artery, где принимаемые и передаваемые сообщения определяются в каждом временном интервале путем моделирования взаимодействия моделей. После этого расчеты данных будут выполнены во внутренних модулях OpenCDA и Middleware обеспечения. Затем моделирование передачи сообщений снова запускается в Artery с обновленными данными для совместного планирования. Наконец, команды управления передаются в CARLA через Artery.

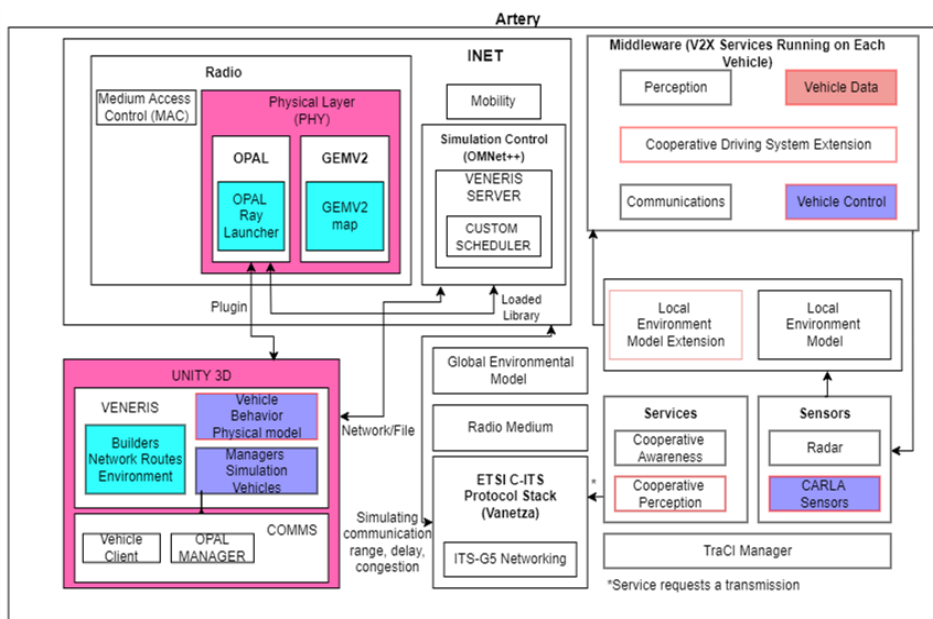


Рис.1. Модуль Artery из архитектуры CAVISE с интегрированным Veneris

На рис.1 показан доработанный модуль Artery из архитектуры CAVISE (добавленные блоки отмечены розовым цветом). Ключевым отличием от предыдущей версии является добавление фреймворка Veneris, который интегрирован в Artery. Veneris состоит из симулятора дорожного движения, реализованного с использованием Unity 3D.

Важно отметить, что все новые компоненты сгруппированы в функциональные модули. Интегрированные инструменты моделирования изображены в белых рамках, а модули, работающие с одними и теми же данными, помечены цветом (бирюзовый для тех, кому требуется синхронизация, и красная рамка для тех, кому требуется расширение).

Модуль Veneris основан на использовании Opal и OMNeT++ и интегрирован в рабочий процесс блока INET. Этот модуль использует Opal для моделирования распространения нескольких лучей, в то время как OMNeT++ служит основой для Opal и GEMV2. Это также необходимо для обеспечения двунаправленного взаимодействия между симулятором трафика и сети. Сервер общего выполнения заменяет стандартный планировщик OMNeT++ внешним планировщиком времени, который увеличивает время моделирования и выполняет события только после получения сообщений синхронизации от внешнего си-

мулятора. Каждое сообщение, полученное от Veneris, содержит время моделирования внешнего симулятора в заголовке, поэтому моделирование продолжается с каждым полученным сообщением.

Заключение

В этой статье рассмотрены особенности моделей распространения сигналов GEMV2 и Opal. В случае GEMV2, которая учитывает реальное расположение объектов, главной особенностью является то, что она обеспечивает хороший компромисс между точностью и масштабируемостью, обеспечивая большой выигрыш в масштабируемости по сравнению с моделями трассировки лучей, обеспечивая при этом достаточную простоту использования. В то время как Opal, благодаря трассировке лучей и работе с графическим процессором, позволяет более детально моделировать распространение беспроводного сигнала [23].

В итоге объединение Opal и GEMV2 в CAVISE даст возможность моделировать распространение сигналов в 3D пространстве, и при этом в зависимости от сценария и поставленных задач, использовать ту модель, которая наиболее оптимальна.

Список литературы

1. Leeson G.W. The growth, ageing and urbanisation of our world //Journal of Population Ageing. – 2018. – Т. 11. – С. 107-115.
2. Papadoulis A., Quddus M., Imprialou M. Evaluating the safety impact of connected and autonomous vehicles on motorways //Accident Analysis & Prevention. – 2019. – Т. 124. – С. 12-22.
3. Stepanyants V., Romanov A. Analysis of Requirements for Next-Generation Complex Urban and Transportation System Simulation //2022 Dynamics of Systems, Mechanisms and Machines (Dynamics). – IEEE, 2022. – С. 1-5.
4. Patella S.M. et al. Carbon Footprint of autonomous vehicles at the urban mobility system level: A traffic simulation-based approach //Transportation Research Part D: Transport and Environment. – 2019. – Т. 74. – С. 189-200.
5. Stepanyants V.G., Romanov A.Y. A Survey of Integrated Simulation Environments for Connected Automated Vehicles: Requirements, Tools, and Architecture //IEEE Intelligent Transportation Systems Magazine. – 2023.
6. CAVISE · GitHub. URL: <https://github.com/CAVISE> (дата обращения: 13.02.2024)
7. Xu R. et al. OpenCDA: an open cooperative driving automation framework integrated with co-simulation //2021 IEEE International Intelligent Transportation Systems Conference (ITSC). – IEEE, 2021. – С. 1155-1162.
8. Riebl R. et al. Artery: Extending veins for VANET applications //2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). – IEEE, 2015. – С. 450-456.
9. Boban M., Barros J., Tonguz O.K. Geometry-based vehicle-to-vehicle channel modeling for large-scale simulation //IEEE Transactions on Vehicular Technology. – 2014. – Т. 63. – №. 9. – С. 4146-4164
10. Egea-Lopez E. et al. Opal: An open source ray-tracing propagation simulator for electromagnetic characterization //Plos one. – 2021. – Т. 16. – №. 11. – С. e0260060.
11. Egea-Lopez E. et al. Vehicular networks simulation with realistic physics //IEEE Access. – 2019. – Т. 7. – С. 44021-44036.
12. Aygun B. et al. Geometry-based propagation modeling and simulation of vehicle-to-infrastructure links //2016 IEEE 83rd Vehicular Technology Conference (VTC Spring). – IEEE, 2016. – С. 1-5.
13. Xiao G. et al. A cooperative offloading game on data recovery for reliable broadcast in VANET //Concurrency and Computation: Practice and Experience. – 2017. – Т. 29. – №. 14. – С. e3938.
14. Breiman L. Random forests //Machine learning. – 2001. – Т. 45. – С. 5-32.
15. Viriyasitavat W. et al. Vehicular communications: Survey and challenges of channel and propagation models //IEEE Vehicular Technology Magazine. – 2015. – Т. 10. – №. 2. – С. 55-66.
16. Parsons J.D., Parsons P. J. D. The mobile radio propagation channel. – New York : Wiley, 2000. – Т. 2.
17. GitHub - NegInfinity/ProjectExodus: Project Exodus - Unity to Unreal scene/project transfer tool. URL: <https://github.com/NegInfinity/ProjectExodus> (дата обращения: 13.02.2024)
18. Utu Plugin - Unity to Unreal Project Converter. URL: <https://alexquevillon.gumroad.com/1/UtuPlugin> (дата обращения: 13.02.2024)
19. Платформа Omniverse для совместной работы над 3D-проектами и моделирования | NVIDIA. URL: <https://www.nvidia.com/ru-ru/omniverse/> (дата обращения: 13.02.2024)
20. GitHub - carla-simulator/carla: Open-source simulator for autonomous driving research. URL: <https://github.com/carla-simulator/carla> (дата обращения: 13.02.2024)
21. Hummel M., van Kooten K. Leveraging nvidia omniverse for in situ visualization //High Performance Computing: ISC High Performance 2019 International Workshops, Frankfurt, Germany, June 16-20, 2019, Revised Selected Papers 34. – Springer International Publishing, 2019. – С. 634-642.
22. INET Framework. URL: <https://inet.omnetpp.org/> (дата обращения: 13.02.2024)
23. Györgyi C. et al. Radio Propagation Digital Twin Aided Multi-Point Transmission With In-Network Dynamic On-Off Switching //IEEE Access. – 2023. – Т. 11. – С. 128843-128858.

ПРОГРАММНЫЕ МОДУЛИ ДЛЯ РАСШИРЕНИЯ СИМУЛЯТОРА NS-3 И ФРЕЙМВОРКА MOREV2X

*Шембель Д.А.¹, Богомолов Р.Д.³, Сивов Г.В.²,
Минибаяева А.И.², Чо Хансли Че¹*

*Национальный исследовательский университет
«Высшая школа экономики»,
¹департамент электронной инженерии,
²департамент компьютерной инженерии,
³департамент прикладной математики
МИЭМ НИУ ВШЭ*

Аннотация

Данная работа направлена на разработку программных модулей для расширения функциональности симулятора сетей ns-3 и расширения фреймворка MoReV2X. В частности, особое внимание уделяется созданию новых модулей мобильности с целью улучшения реалистичности симуляций, а также разработке инструментов для постобработки результатов моделирования, что позволит исследователям анализировать и визуализировать ключевые метрики производительности сети. Эти улучшения позволят более точно моделировать прямую связь между автономными транспортными средствами в сетях 5G NR-V2X.

Введение

Прямая межавтомобильная коммуникация без необходимости охвата базовой станцией (режим № 2, sidelink) вместе с механизмом полупостоянного планирования для автономного распределения радиоресурсов между транспортными средствами представляет собой многообещающий подход для поддержки периодической передачи в

5G NR-V2X, что является основой для кооперативных приложений автоматизированного вождения. [1]

Поскольку в данный момент отсутствует доступная экспериментальная база для проведения натуральных экспериментов с реальными транспортными устройствами, исследователи широко используют средства имитационного моделирования. В настоящее время существует ns-3 (Network Simulator 3), дискретно-событийный симулятор сетей, который предоставляет гибкую платформу для моделирования и анализа различных сетевых протоколов, технологий и сценариев.

Также существует расширение над ns-3 в виде MoReV2X [2], который представляет собой инструмент для моделирования и анализа сетей автономных автомобилей и других участников дорожного движения на базе сетей связи нового поколения (5G и выше). MoReV2X предоставляет возможность моделирования коммуникаций между различными типами устройств, такими как автомобили, инфраструктура дорог, пешеходы и другие. MoReV2X специализируется на моделировании режима прямой связи между автомобилями с использованием 5G.

Однако несмотря на то, что MoReV2X предоставляет широкие возможности по моделированию сетевой коммуникации между транспортными средствами, MoReV2X обладает некоторыми недостатками, среди которых можно выделить:

1. Недостаточный уровень реализма в моделях мобильности транспортных средств: используемая модель мобильности с постоянной скоростью не является реалистичной, что может привести к искаженным результатам при симуляции мобильности.
2. Отсутствие функционала постобработки: В MoReV2X отсутствует открытый функционал постобработки полученных результатов во время моделирования, что ограничивает возможности анализа.
3. Ограниченные возможности анализа: Недостаточное количество инструментов для анализа данных и выведения ключевых выводов из симуляций может затруднять процесс исследования и принятия решений на основе полученных результатов.

Перечисленные проблемы позволяют производить ограниченный перечень исследований прямой связи между автономными автомобилями.

В рамках данной работы мы описываем результаты разработки программных модулей для расширения функционала симулятора MoReV2X в соответствии с архитектурой, представленной на рис.1. В рамках нашей работы мы предлагаем программный интерфейс для интеграции файлов реалистичной мобильности в MoReV2X, а также программные модули для постобработки и визуализации результатов симуляций, проведенных в MoReV2X.

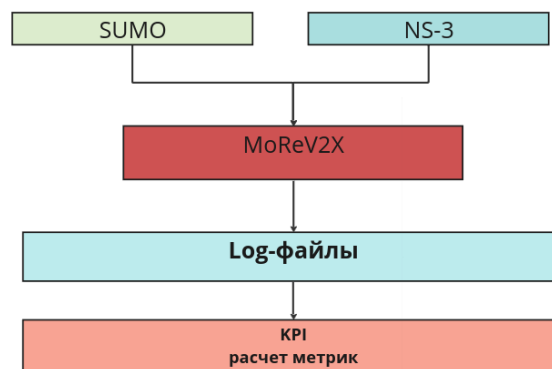


Рис.1. Архитектура разрабатываемого решения

Программный интерфейс для интеграции trace-файлов SUMO в ns-3 MoReV2X

Создание новых модулей мобильности позволяет более детально исследовать новые гипотезы применительно к сценариям движения автомобилей по шоссе и в городской среде. Для проведения симуляций с реалистичной мобильностью автомобилей необходимо использовать профессиональные симуляторы, позволяющие генерировать реалистичные транспортные потоки в виде выходных trace-файлов.

Для этих целей широко используется симулятор SUMO (Simulation of Urban Mobility) [10]. Однако, в настоящее время интеграция trace-файлов SUMO в ns-3 MoReV2X не произведена, что не позволяет оценить работу алгоритмов распределения ресурсов в реалистичных сценариях движения потока автомобилей.

В нашей работе был разработан программный интерфейс, позволяющий использовать результаты SUMO симуляций внутри ns-3 MoReV2X. Внутри SUMO происходит расчет позиций, скоростей, ускорений автомобилей по времени, после чего на выходе пользователь получает .xml файлы с результатами симуляций. Разработанный программный интерфейс переводит trace-файлы SUMO в формат .csv, с которым работает MoReV2X, после чего полученные координаты используются для создания нод, на которые могут быть наложены сетевые точки, для моделирования сетевого взаимодействия движущихся узлов. Архитектура описанного интерфейса представлена на рис.2.

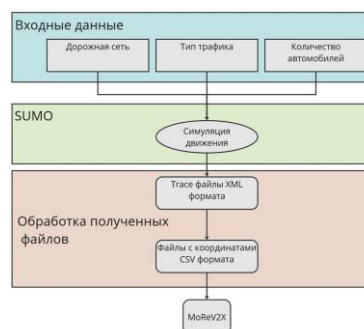


Рис.2. Принцип работы программного интерфейса

Программный модуль для пост-обработки данных симуляций

В результате моделирования с использованием MoReV2X формируются датасеты в виде Log-файлов.

В результате нашей работы был разработан программный модуль, позволяющий обрабатывать полученные после моделирования датасеты и рассчитывать ключевые метрики производительности сети автономных автомобилей.

Основные ключевые метрики и их значение:

1. Возраст информации (Age of Information, AoI) - возраст (время) наиболее недавней единицы информации, полученной от ресурса, посылающего обновления.

$$E[AoI] = \frac{\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T AoI(t) dt}{\lim_{N \rightarrow \infty} \frac{\sum_{k=1}^N \frac{1}{2} Y_k^2}{\sum_{k=1}^N Y_k}}$$

где Y_k равняется времени между двумя последовательно полученными пакетами для каждой пары узлов.

2. Загруженность канала связи (Channel Busy Ratio, CBR) - отношение количества загруженных подканалов к общему количеству подканалов; $CBR = n(3)$, где n - доля занятых подканалов в $[n - 100 \cdot 2^{\mu}, n - 1]$ предыдущих интервалов. Подканал помечается как занятый, если показатель уровня принимаемого сигнала (RSSI) превышает предварительно настроенное пороговое значение.

3. Коэффициент успешной доставки пакетов (Packet Delivery Ratio, PDR) - отношение успешно переданных пакетов к общему количеству отправленных пакетов:

$$PDR = \frac{\sum_{i=1}^M Y_i}{\sum_{i=1}^M Z_i} \quad (4)$$

где Y - количество автомобилей в регионе, которые успешно принимают пакет, а Z - общее количество автомобилей в том же регионе, M - общее количество пакетов, которое генерируется во время моделирования, пределы региона находятся на расстоянии от передающего автомобиля в диапазоне $[a; b]$, где

$$a = i \cdot 50 \text{ м}, b = (i + 1) \cdot 50 \text{ м}, i = 0; 18 \quad (5)$$

Коэффициент потерь из-за затухания сигнала (Propagation Loss Ratio, PLR) - соотношение количества пакетов, потерянных из-за плохих условий распространения (недостаточное отношение сигнал/шум (SNR)) в диапазоне расстояний $[a; b]$ (см. 168

формулу (5)) к общему количеству пакетов, переданных в этом диапазоне:

$$PLR = \frac{N_{PL}}{N_{PL} + N_{CL} + N_{SR}} \quad (6)$$

где N_{PL} — это количество пакетов, полученных неправильно в одном и том же диапазоне из-за плохих условий распространения (т. е. количество пакетов, которые не столкнулись, но получили недостаточное SNR), N_{CL} — это количество пакетов в $[a, b]$ (см. формулу (5)) диапазон расстояний, которые были потеряны в результате коллизии и не могли быть успешно восстановлены из-за недостаточного отношения сигнал/шум+интерференция, N_{SR} — это количество правильно полученных пакетов в том же диапазоне.

Коэффициент потерь из-за коллизий пакетов (Collision Loss Ratio, CLR) – по аналогии с PLR соотношение количество пакетов, потерянных из-за коллизий в диапазоне расстояний от $[a, b]$ (см. формулу (5)) к общему количеству пакетов, переданных в этом диапазоне:

$$CLR = \frac{N_{CL}}{N_{PL} + N_{CL} + N_{SR}} \quad (7)$$

Программный модуль визуализации данных симуляции

На следующем этапе пост-обработки данных необходимо провести визуализацию для анализа зависимостей ключевых метрики производительности сети от параметров симуляции. Модуль визуализации получает на вход обработанные файлы, в которых содержится информация о рассчитанных метриках и их зависимости от дистанции между автомобилями, коэффициента персистентности и других параметров, важных для алгоритмов распределения радиоресурсов в сети.

После выбора интересующих зависимостей, которые необходимо визуализировать, и выбранных настроек визуализации (ширина, высота графика, границы осей, цвет графика, маркеры), формируются графики функций графики и сохраняются в .eps файлах.

На данный момент модуль предоставляет следующие зависимости:

1. Ключевые метрики производительности как функция расстояния между узлами.
2. Ключевые метрики производительности как функция вероятности персистентности узлов.
3. Ключевые метрики производительности как функция мощности сигнала передатчика.

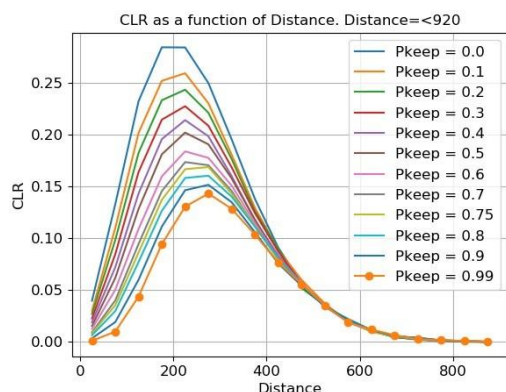


Рис.3. Визуализация зависимости коэффициента потерь из-за коллизий пакетов от расстояния между узлами при различных значения вероятности персистенции узлов

На рис.3 представлен результат работы программного модуля визуализации в виде построенных графиков зависимостей от параметров симуляции. Результаты симуляции, использованные для построения графиков, были получены от авторов статьи [4].

Заключение

Результаты данной работы представляют значительный вклад в улучшение функциональности симулятора сетей ns-3 и расширение фреймворка MoReV2X. Особое внимание уделяется более реалистичным симуляциям и разработке инструментов для постобработки результатов моделирования. Эти усовершенствования позволят исследователям более точно анализировать ключевые метрики производительности сети и моделировать прямую связь между автономными транспортными средствами в сетях 5G NR-V2X.

В рамках дальнейшей работы над расширением возможностей симулятора планируется проведение широкомасштабного тестирования разработанных модулей по результатам моделирования в сценариях движения транспортных средств по шоссе с наличием скрытых узлов и формированием групп транспортных средств с различной плотностью на различных участках шоссе.

Список литературы

1. Q. Delooz et al., "Analysis and evaluation of information redundancy mitigation for v2x collective perception", IEEE Access, vol. 10, pp. 47 076-47 093, 2022.
2. L. Lusvarghi and M. L. Merani, "MoReV2X - A New Radio Vehicular Communication Module for ns-3," 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Norman, OK, USA, 2021, pp. 1-7, doi: 10.1109/VTC2021-Fall52928.2021.9625478.
3. P. A. Lopez et al., "Microscopic Traffic Simulation using SUMO," 2018 21st International Confer-

ence on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 2018, pp. 2575-2582, doi: 10.1109/ITSC.2018.8569938.

4. A. Rolich, I. Turcanu, A. Vinel and A. Baiocchi, "Impact of Persistence on the Age of Information in 5G NR-V2X Sidelink Communications," 2023 21st Mediterranean Communication and Computer Networking Conference (MedComNet), Island of Ponza, Italy, 2023, pp. 15-24, doi: 10.1109/MedComNet58619.2023.10168874

ОБЕСПЕЧЕНИЕ НАБЛЮДАЕМОСТИ СИСТЕМ НА ОСНОВЕ МИКРОСЕРВИСНОЙ АРХИТЕКТУРЫ

Курилович П.А.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В работе рассматриваются способы наблюдения за приложениями, основанными на микросервисной архитектуре. Описываются возможные данные, которые можно снимать во время работы системы, а также направления для их анализа.

Введение

С развитием высоконагруженных систем и увеличением требований к производительности, микросервисная архитектура становится все более распространенной. Эффективное управление и мониторинг таких систем требует особого подхода. В этом контексте важную роль играет наблюдаемость и его анализ, то есть способность понимать внутреннее состояние системы на основе внешних наблюдений. Цель данного исследования - рассмотреть принципы и методы обеспечения наблюдаемости в микросервисной архитектуре для обеспечения высокой производительности и устойчивости системы при высоких нагрузках, а также обозначить направления для его анализа.

Актуальность

По статистике Google Trends популярность микросервисной архитектуры растёт который год (рис. 1). Развитие микросервисов отчасти было обусловлено пропагандой их преимуществ такими технологическими гигантами, как Netflix, Amazon, Twitter (X), Google и Uber.



Рис.1. Популярность микросервисной архитектуры, начиная с 2004 года

В связи с нарастающими решениями внедрения именно микросервисной архитектуры для решения бизнес задач, возрастает и количество проблем которые с ними связаны. Одна из таких проблем - это анализ и мониторинг работоспособности всех микросервисов в системе. В настоящее время существуют системы с сотнями и тысячами микросервисов, взаимодействующих друг с другом, и компаниям стало очень сложно принимать бизнес-решения для таких больших и сложных систем. Часто это может вызывать проблемы при обнаружении связей между службами, а также проблемы с производительностью.

В различной литературе рассматриваются вопросы обнаружения данных проблем, а также плохих показателей в микросервисной архитектуре. Например, Тайби и др. [1] выявили повторяющиеся проблемы в микросервисах. Уокер и др. [2], а также Пигаццини и др. [3] обратили внимание на эти проблемы и предложили способы их решений. Основной подход включал в себя статический анализ или комбинацию статического и динамического анализа. Основная задача состоит в том, чтобы восстановить целостную системную картину. Выяснить способы, которые позволяют программистам понять, как система делится на конкретные микросервисы и как они взаимодействуют и зависят друг от друга.

Телеметрия

В разработке программного обеспечения данные телеметрии относятся к сбору данных из программного обеспечения и систем, которые указывают состояние источника. Телеметрия – один из ключевых факторов повышения наблюдаемости сложной системы. Johnson и др. [4] определили телеметрию программного продукта как метод определения, сбора и анализа метрик программного обеспечения, который имеет пять характеристик, перечисленных ниже:

1. Данные должны собираться автоматически, без ручного вмешательства человека.
2. Каждому событию в данных должна быть присвоена временная метка.
3. Каждый участник проекта имеет постоянные и немедленный доступ к данным.
4. Анализ телеметрии должен быть ценным, даже если в нем отсутствуют полные данные за весь период существования проекта.
5. Анализ телеметрии отображает текущее состояние проекта и его развитие с течением времени.

OpenTelemetry [5] - это стандартизация процесса сбора телеметрических данных. Разработчикам сложно переключаться на разные инструменты и

адаптироваться к новому инструменту из-за различной стандартизации, существовавшей до OpenTelemetry. OpenTelemetry предоставляет независимый от поставщика API для отправки данных телеметрии на серверную часть и набор библиотек для конкретного языка для внедрения инструментов сбора телеметрии в код и доставки данных на одну из поддерживаемых серверных частей.

Данные телеметрии можно разделить на три основных формата [5][6][7]:

1. Логи.
2. Трассировки (трейсы).
3. Метрики.

Лог - это текстовая запись с меткой времени, структурированная (рекомендуется) или неструктурированная, содержащая метаданные. Из всех сигналов телеметрии логи имеют самое большое значение. На данный момент для логов существует стандарт RFC-3164, который описывает правила и формат логирования [8].

Трассировки, также называемые операциями, связывают этапы одного запроса с несколькими вызовами внутри и между микросервисами. Они могут обеспечить структурированное наблюдение за взаимодействием компонентов системы. Трассировка может начинаться на ранних этапах процесса запроса, например, в пользовательском интерфейсе приложения, и может распространяться через сетевые службы по сети микросервисов, которые обрабатывают запрос. В OpenTelemetry трейсы неявно определяются их промежутками/интервалами (spans). Трассировки указывают, как запросы распространяются через наши службы и другие компоненты. В частности, трассировку можно рассматривать как ориентированный ациклический граф (DAG) промежутков, где ребра между промежутками определяются как отношения родитель/потомок [5] (рис.2).

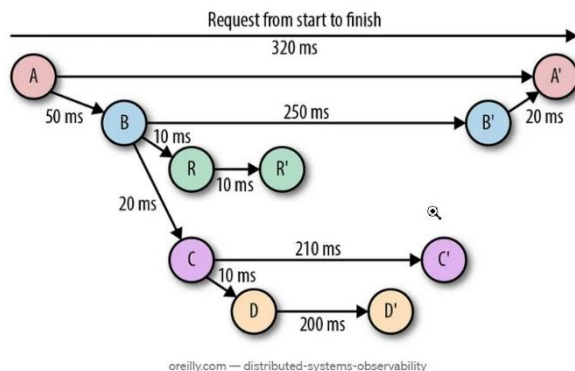


Рис.2. Граф зависимостей сервисов при совершении запроса [5]

На основе трейсов можно построить график зависимости сервисов (SDG) [9]. Он позволяет отследить связи между всеми микросервисами в системе.

Метрика - это некоторое измерение сервиса, полученное во время выполнения. Момент регистрации измерений известен как метрическое событие. Оно состоит из самого измерения, времени получения и связанных с ним метаданных.

Чтобы сделать метрику полезной, мы отслеживаем ее состояние, обычно записывая точки данных с течением времени. Эти точки данных называются наблюдениями. Наблюдение состоит из значения, временной метки и иногда ряда свойств, описывающих наблюдение, таких как источник или теги. Набор наблюдений называется временным рядом.

Анализ микросервисов

За счет логов можно не только отслеживать ошибки или заниматься отладкой, но и получать полезную информацию в ходе цепочки вызовов между сервисами, особенно если применять структурное логирование и трейсинг.

На основе трейсинга телеметрии возможен анализ всей системы по ряду направлений, а именно:

А. Циклические зависимости сервисов в системе: Цикл в графе, состоящий из одного, двух и более микросервисов. Несмотря на то, что запросы не зацикливаются, поскольку не вызываются одни и те же конечные точки, это все равно плохая практика, которая может привести к ухудшению архитектуры. Метрика SIY была определена Рудом и другими [10] как количество пар, зависящих друг от друга. Мы можем просмотреть каждую пару (a, b), чтобы увидеть, существует ли путь от сервиса «a» к сервису «b» и наоборот. Метрика SIY - это количество таких пар.

В. Узкое место (bottleneck) системы: Один или несколько сервисов, на которые приходится большая часть всей нагрузки в системе. Этот антипаттерн появляется, когда у сервиса большое количество запросов, и сервис становится единой точкой входа/отказа [1].

С. Тепловая карта (heatmap) на основе SDG в сочетании с тепловой картой (количество пересечений сервиса с другим) может помочь эффективно масштабировать систему с точки зрения потребления ресурсов. Например, с помощью этого способа можно определить наиболее часто используемые сервисы и предпринять необходимые шаги для улучшения работоспособности (например, создать больше реплика сервисов).

Д. Общие ресурсы и версии API [11]: На основе SDG можно выяснить, используются ли одни и те же ресурсы (например, БД) несколькими сервисами одновременно. Также это позволяет отследить, имеет ли сервис несколько версий API (на основе endpoint'ов каждого сервиса).

Е. Абсолютная зависимость от сервиса (ADS) [11]: Можно найти скалярную величину, показывающую сколько сервисов зависят от одного конкретного (количество ребер на графе). Если это число велико, это повод обратить внимание на такой микросервис, возможно имеет смысл «разнести» его на несколько других.

На основе метрик имеется возможность отслеживать широкий спектр полезной информации, вот некоторые из них:

1. Загруженность сервиса: количество запросов на API сервиса по различным протоколам (например, HTTP/gRPC) за единицу времени (как правило за секунду) (RPS) [12]. Это позволяет определить наиболее уязвимые и важные узлы во всей системе.

2. Разброс скорости ответа на API сервиса: гистограмма распределения времени обработки запросов или перцентили [5]. Благодаря этой информации можно обнаружить проблемные места в конкретном сервисе на конкретном endpoint'е API или же в целом определить сервисы, которые не удовлетворяют скорости обработки запроса.

3. Информация о ресурсах: снятие показателей об утилизации CPU и расхода памяти RAM, количестве потоков, размер свободной памяти на диске. Все это позволяет заранее определять возможную нехватку ресурсов или обращать внимание на явные проблемы с производительностью.

Универсальный дашборд

Исходя из направлений анализа микросервисов предлагается сделать универсальный дашборд в Grafana с метриками, который можно импортировать и использовать для любого сервиса. Для этого нужно лишь переопределить глобальные переменные. На данный момент ни один продукт не предлагает такой возможности, а лишь предоставляет инструменты для реализации самостоятельно.

Например, следуя пункту о необходимости знать загруженность сервиса уместно снимать показания RPS и общую задержку (Latency) сервиса. Стоит отметить, что показания снимаются не только на весь трафик, но и по каждому методу отдельно. Для каждой метрики рассчитываются максимальное, минимальное и среднее значения за время наблюдения. Пример таких метрик для сервиса представлен ниже (рис.3).

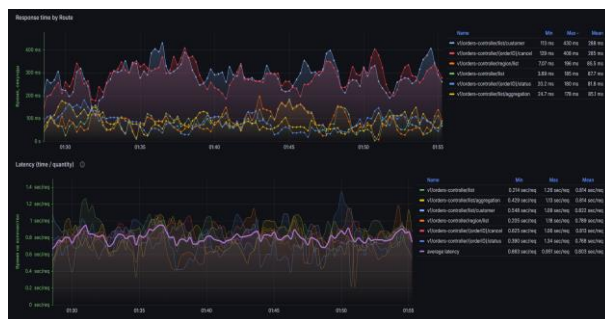


Рис.3. Графики нагрузки на сервис (количество запросов в секунду и задержка)

Также крайне важным является оценка скорости обработки запроса в целом и разбиение всех запросов на промежутки или же квантили. Они могут устанавливаться как самостоятельно, так и автоматически. По своей сути это является гистограммой, однако форма представления может быть разной. Например, глядя на тепловую карту можно моментально оценить производительность сервиса и его динамику с течением времени. В тоже время обычная столбчатая диаграмма позволяет оценить конкретное число запросов в промежутке. Пример таких метрик для сервиса представлен ниже (рис.4).



Рис. 4. Тепловая карта и распределение длительности запроса по промежуткам

Заключение

Наблюдаемость систем, основанных на микросервисной архитектуре, является очень важным пунктом при разработке в последние годы. Не существует международных стандартов по мониторингу микросервисных систем, однако существуют рекомендации популярных компаний, которым пытаются следовать все другие [13].

В данной статье рассмотрены данные, снятие которых позволяет проводить анализ микросервисов, возможные направления для анализа, их польза при построении системы и пример реализации этих метрик в Grafana.

Унификация метрик, способов логирования и трейсинга являются предметом дальнейших исследований.

Список литературы

1. Taibi D., Lenarduzzi V. On the definition of microservice bad smells //IEEE software. – 2018. – Т. 35. – №. 3. – С. 56-62.
2. Walker A., Das D., Cerny T. Automated code-smell detection in microservices through static analysis: A case study //Applied Sciences. – 2020. – Т. 10. – №. 21. – С. 7800. [Online] Available: <https://www.mdpi.com/2076-3417/10/21/7800>
3. Pigazzini I. et al. Towards microservice smells detection //Proceedings of the 3rd International Conference on Technical Debt. – 2020. – С. 92-97. [Online] Available: <https://doi.org/10.1145/3387906.3388625>
4. Johnson P. M. et al. Improving software development management through software project telemetry //IEEE software. – 2005. – Т. 22. – №. 4. – С. 76-85.
5. «Opentelemetry: High-quality, ubiquitous, and portable telemetry to enable effective observability», 2024 г. [Электронный ресурс] URL. Доступно: <https://opentelemetry.io/> (Дата обращения: 01.01.2024)
6. S. Karumuri, F. Solleza, S. Zdonik, and N. Tatbul, "Towards observability data management at scale," ACM SIGMOD Record, vol. 49, no. 4, pp. 18–23, 2021.
7. Picoreti R. et al. Multilevel observability in cloud orchestration //2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). – IEEE, 2018. – С. 776-784.

8. Стандарт RFC-3164 [Электронный ресурс] URL: <https://datatracker.ietf.org/doc/html/rfc3164.html> (Дата обращения: 12.01.2024)

9. Al Maruf A. et al. Using microservice telemetry data for system dynamic analysis //2022 IEEE International Conference on Service-Oriented System Engineering (SOSE). – IEEE, 2022. – С. 29-38.

10. Rud D., Schmietendorf A., Dumke R. R. Product metrics for service-oriented infrastructures //IWSM/MetriKon. – 2006. – С. 161-174.

11. Palma F., Mohay N. A study on the taxonomy of service antipatterns //2015 IEEE 2nd International Workshop on Patterns Promotion and Anti-patterns Prevention (PPAP). – IEEE, 2015. – С. 5-8.

12. Engel T. et al. Evaluation of microservice architectures: A metric and tool-based approach //Information Systems in the Big Data Era: CAiSE Forum 2018, Tallinn, Estonia, June 11-15, 2018, Proceedings 30. – Springer International Publishing, 2018. – С. 74-89.

13. "Four Golden Signals" [Online] Available: <https://sre.google/sre-book/table-of-contents/>

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ОПТИМИЗАЦИЙ КОМПИЛЯТОРОВ CLANG И GCC НА ПРИМЕРЕ КОМПИЛЯЦИИ ВЫСОКОНАГРУЖЕННОЙ СИСТЕМЫ

Арутюнян А.Р.¹, Осинцев Н.И.²,
Омаров М.Т.², Нестеренко А.А.³

Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент прикладной математики,

²департамент компьютерной инженерии,

³департамент электронной инженерии

МИЭМ НИУ ВШЭ

Аннотация

Данная статья осуществляет сравнительный анализ оптимизационных возможностей компиляторов Clang и GCC.

Несмотря на широкое распространение обоих компиляторов, их различия в оптимизации крупномасштабных проектов остаются мало исследованными. Цель работы заключается в сравнении архитектур компиляторов и интерпретации результатов компиляции проекта VK.com.

В работе был проведен анализ документации по архитектуре компиляторов, подготовлена среда для компиляции проекта в средах Clang и GCC с различными опциями оптимизации и проанализированы показатели производительности скомпилированного кода.

Введение

Компилятор [1] – инструмент перевода высокоуровневого кода в машинные инструкции. На рынке наиболее популярными решениями для работы с семейством C-языков выступают GCC (GNU Compiler Collection) [2] и Clang [3].

GCC является проектом с длительной историей и считается наиболее популярным для разработки компилятором. Для поддержки множества языков программирования все оптимизации GCC проводит над общим независимым от языка программирования представлением, часть этих оптимизаций также не зависит от целевого устройства сборки, что сделало иерархии внутренних представлений сложными, а структуру оптимизаций монолитной.

Clang – относительно новый компилятор C-подобных языков, разработанный на базе LLVM-фреймворка и обладающий модульной архитектурой. Clang за последние годы благодаря своей открытой лицензии и активной поддержке ведущими компаниями IT-индустрии приобрел огромную популярность в сообществе разработчиков.

Каждый из компиляторов состоит из фронтенд (работа с высокоуровневым представлением) и бэк-енд (работа с языко-независимым представлением) частей. Основное различие заключается во временном представлении, которое предоставляет Clang: оно значительно проще, чем GCC-представление. Осуществление оптимизаций в Clang проводится иначе, чем в GCC и не имеет четкой последовательности действий. При этом модульная архитектура Clang'a позволяет легко дорабатывать компилятор.

Ниже будет более подробно рассмотрена архитектура каждого из проектов с выделением их особенностей.

Clang и LLVM

Сравнение архитектур Clang и GCC можно наблюдать на рис.1.

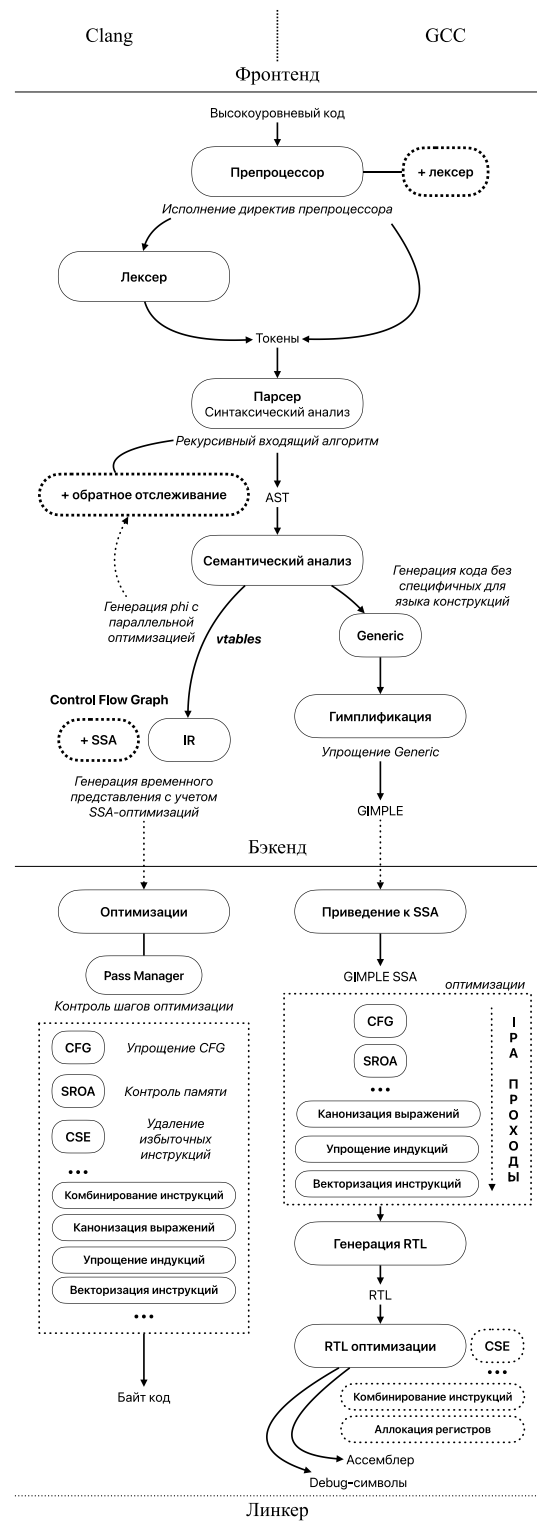


Рис.1. Архитектуры компиляторов Clang и GCC

Как видно из рисунка, оба компилятора используют для генерации синтаксического дерева AST свои рекурсивные сходящие парсеры. Однако Clang отличается дополнительным функционалом отслеживания (backtracing) [4], которое активно используется при генерации временного представления IR (Intermediate Representation).

Другим более серьезным отличием служат принципы построения языко-независимых представле-

ний: на стороне Clang'a используется IR LLVM, которое отличается простотой конструкции и читабельностью. Стоит отметить, что генерация IR производится напрямую в SSA-форму [5], что значительно сокращает код и ускоряет дальнейшие оптимизации. В GCC же фронтенд исключает из AST высокоуровневые синтаксические конструкции для создания GENERIC представления. Далее, GENERIC дерево преобразовывают с целью сохранения только последовательного порядка выполнения и безусловных/условных переходов. Этот этап формирует GIMPLE – конечный результат работы фронтенда. На этапе бэкенда GIMPLE преобразуется в форму SSA, после чего над ним проводятся платформо-независимые оптимизации и IPA проходы (внутри-процедурные оптимизации) для LTO (оптимизации во время линковки) [6].

LLVM, выступающий бэкендом компилятора Clang, умело использует use-def и def-use [7] цепочки, которые формируются за счет SSA-представления, такие же механизмы отслеживания переменных используются и в GCC.

Наиболее интересной для анализа компиляторов выступает стадия оптимизаций. Clang использует свой менеджер проходов (Pass Manager), который способен сам определять очередность вызовов определенных этапов оптимизации. Среди них наиболее важными можно назвать:

- упрощение Control Flow Graph (CFG): удаление пустых блоков, объединение блоков;
- контроль памяти: приведение выделение памяти к виртуальным регистрам SSA;
- упрощение индукций: канонизация управляющих блоков циклов, их оптимизация;
- векторизация инструкций: распараллеливание вычислений схожих операций;
- и др.

В то же время GCC не обладает таким механизмом распределения этапов оптимизаций [8]: выделена определенная последовательность этапов, которые компилятор может применить для улучшения представления GIMPLE, а затем RTL.

Умение определять очередность оптимизационных этапов значительно влияет на работоспособность больших проектов и делает их более стабильными, о чем будет сказано позже.

Сравнение оптимизаций

Рассмотрим на примере способности GCC и Clang оптимизировать исходный код. В качестве примера будем использовать программу, представленную ниже (рис.2).

```
int dummy_mul_by_1024(int x) {
    int dummy_1024 = 1;
    for (int i = 0; i < 10; i++) {
        dummy_1024 = (-2) * dummy_1024;
    }
    return dummy_1024 * x;
}
```

Рис.2. Код функции

Функция `dummy_mul_by_1024` производит умножение числа, переданного в качестве аргумента, на 1024, и возвращает результат. Очевидно, что представленная реализация не оптимальна – она 10 раз умножает число на -2 вместо одноразового умножения на 1024. Однако, на таком примере можно заметить отличия в оптимизациях, которые производят компиляторы.

Запустим компиляцию программы из примера с помощью обоих рассматриваемых компиляторов и выведем сгенерированный ими ассемблерный код. Компилировать будем с одинаковым набором флагов компиляции, а именно с флагом `-Os`, который указывает компилятору использовать все оптимизации уровня `-O2`, не увеличивающие объем исполняемого файла.

Рассмотрим ассемблерный код, сгенерированный компилятором GCC версии 9.4.0:

```
movl $10, %edx
movl $1, %eax
.L2:
imull $-2, %eax, %eax
decl %edx
jne .L2
imull %edi, %eax
ret
```

Рис.3. Ассемблерный код от GCC

Код в блоке L2 полностью соответствует исходному: значение в регистре EAX умножается на -2, значение в регистре EDX уменьшается на единицу, после чего управление переходит обратно метке L2, если значение в регистре EDX не равно 0.

Теперь рассмотрим ассемблерный код, сгенерированный компилятором Clang версии 8.0.1 (рис.4).

```
movl %edi, %eax
shll $10, %eax
retq
```

Рис.4. Ассемблерный код от Clang

Сразу заметим, что количество вызываемых инструкций стало меньше. Это произошло за счет того, что Clang в `compile-time` посчитал, во сколько раз умножается аргумент функции за 10 итераций цикла. Инструкция `shll $10, %eax` сдвигает значение регистра EAX на 10 влево, что эквивалентно умножению на 1024. Таким образом, асимптотическая сложность умножения стала равна $O(1)$ вместо $O(N)$, где N – количество итераций цикла.

Разберем, при помощи каких оптимизаций Clang упростил исходный код. Для этого выведем IR-представление исходного кода после каждой из проведенных оптимизаций. Ниже представлена последовательность оптимизаций и результатов преобразований, результатом которых послужило упрощение цикла до операции побитового сдвига (рис.5).

```

*** IR Dump Before Induction Variable Simplification ***
%6 = phi i32 [ 0, %1 ], [ %9, %5 ]
%7 = phi i32 [ 1, %1 ], [ %8, %5 ]
%8 = mul nsw i32 %7, -2
%9 = add nuw nsw i32 %6, 1
%10 = icmp ult i32 %9, 10
br i1 %10, label %5, label %2

%3 = phi i32 [ %8, %5 ]
%4 = mul nsw i32 %3, %0
ret i32 %4
*** IR Dump Before Delete dead loops ***
; Preheader:
%5 = phi i32 [ 0, %1 ], [ %6, %4 ]
%6 = add nuw nsw i32 %5, 1
%7 = icmp ne i32 %6, 10
br i1 %7, label %4, label %2

%3 = mul nsw i32 1024, %0
ret i32 %3
*** IR Dump Before MergedLoadStoreMotion ***
%3 = mul nsw i32 1024, %0
ret i32 %3

```

Рис.5. IR-код этапов оптимизации Clang.

Из этих данных можно сделать вывод о том, что оптимизация «Induction Variable Simplification» упростила тело цикла и вынесла умножение на 1024 в отдельную инструкцию.

Таким образом, в некоторых частных случаях Clang при применении одинаковых флагов компиляции лучше оптимизирует логику исходного кода, чем GCC.

Сравнение на монолите VK

VK.com написан на PHP, который транспируется через КНРП в обширный проект на языке С++, состоящий из порядка 200 тысяч файлов формата .cpp и h. В настоящее время на продакшн-серверах VK.com компилируется через GCC с использованием флага оптимизации -Os.

В рамках данной работы была подготовлена следующая среда для анализа компиляций проекта:

- сервер для сборки бинарных файлов:
OS: Debian GNU/Linux 11 (bullseye)
CPU: Intel Xeon Gold 62238R @ 4.00GHZ
RAM: 512 GB;
- несколько серверов для сборки метрик:
OS: Debian GNU/Linux 11 (bullseye)
CPU: Intel Xeon Gold 6338 @ 2.00GHz
RAM: 128 Gb.

Все эксперименты проводились на реальном продуктивном трафике.

В ходе первого эксперимента было сопоставлено время сборки и линковки, а также размеры бинарных файлов при компиляции с GCC версии 10.2.1-6 и с Clang версии 16.0.6, используя оптимизации -Os, с которыми происходит сборка проекта на продакшн.

В результате были получены следующие результаты (таблица Таблица 1):

Таблица 1. Результаты компиляции VK.com

| | GCC ¹ dwarf4 /32 | GCC ¹ dwarf5 /32 | GCC ² dwarf5 /64 | Clang dwarf5 /32 | Clang dwarf5 /64 |
|--|-----------------------------------|-----------------------------------|-----------------------------------|------------------------|------------------------|
| | | | | | |

| | | | | | |
|----------------------------|--------|--------|--------|--------|--------|
| Компиляция (с) | 2405.2 | 2434.4 | 3107.3 | 2894.4 | 2911.9 |
| Линковка (с) | 15.9 | 16.8 | 19 | 13.7 | 13.5 |
| base- debug.bin (Гб) | 1.2 | 1.2 | 1.9 | 1.4 | 2.1 |
| base.debug (Гб) | 5.6 | 6 | 7.4 | 4 | 4.7 |

Примечание. «base-debug.bin» обозначает размер сборки без дебаг-символов, «base.debug» – размер вынесенных в отдельный файл дебаг-символов. GCC¹ – 10.2.1-6; GCC² – 12.3.1.

В ходе первого эксперимента проект был собран с GCC 10.2.1-6, который по умолчанию использует dwarf4 с 32 битной разрядностью, и сравнивался с Clang dwarf5/32. Основные флаги сборки (рис.6):

GCC:

```
-Os -march=sandybridge -Wa,--compress-debug-sections -Wno-unused-but-set-variable -std=gnu++17
```

Clang:

```
-Os -march=sandybridge -fbracket-depth=350 -Wno-unused-but-set-variable -Wno-unused-but-set-variable -Wno-unused-value -Wno-unused-private-field -Wa,--compress-debug-sections -std=gnu++17
```

ld:

```
-wl,--build-id=uuid -wl,-z,relro -wl,--as-needed -no-pie
```

Рис.6. Флаги компиляции и линковки

Результаты первого теста:

- размер дебаг-секций GCC больше, чем у Clang на 1.6 Гб;
- размер бинарного файла без дебаг-секций у GCC меньше на 0.2 Гб;
- время компиляции у Clang больше на 489 с.

Дебаг-секции связывают машинные инструкции с высокоуровневым кодом. Меньшее количество дебаг-символов говорит о лучшей структуре оптимизированного кода, который позволяет использовать короткие адреса для ссылки на строку кода. Clang позволяет экономить сеть и память на релизах. На продакшн VK.com собирается через посс, распределяющий нагрузку компиляции по серверам, поэтому задержки Clang сократятся и разница не будет заметна.

Также был проанализирован трафик VK.com на следующих метриках:

- kphp_memory_script_usage – используемая в работе память (рис.7);

- `kphr_cpu_usage` – использование процессора (рис.8);
- `kphr_request_time` – время ответа на запрос (рис.9).

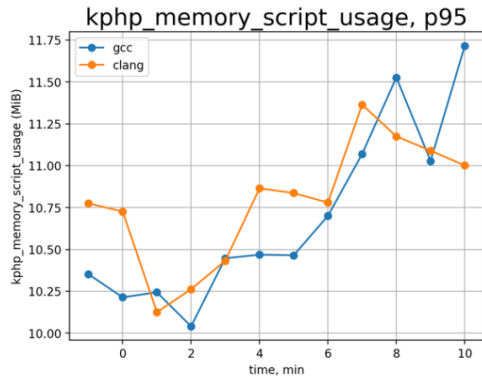


Рис.7. Потребление памяти

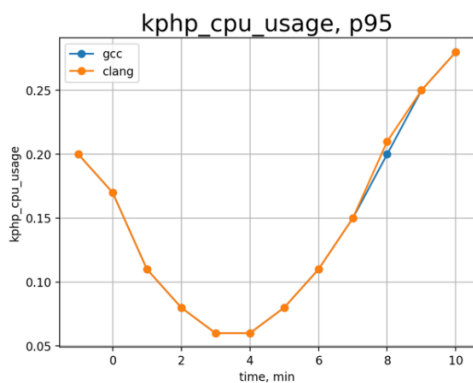


Рис.8. Использование процессора

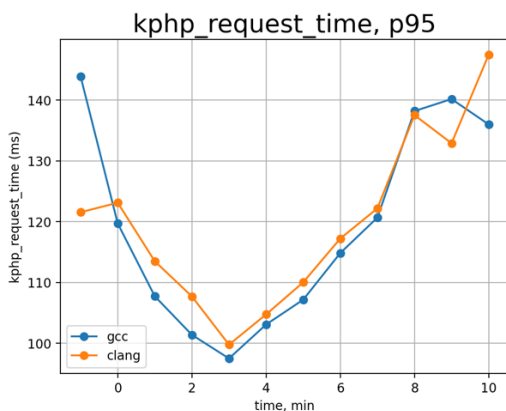


Рис.9. Время ответа на запросы

По графикам можно сделать вывод, что Clang и GCC не имеют существенных различий в производительности исполняемого файла.

Для полноты эксперимента были проведены тесты с использованием GCC с `dwarf5/32`, однако из таблицы Таблица 1 видно, что размер бинарного файла на стороне GCC стал еще больше, чем у Clang.

Наконец, для учета проблемы переполнения дебаг-секций, которая может возникнуть на больших проектах, был использован `dwarf5/64` с GCC 12.3.1. В результате время компиляции у GCC значительно выросло и превысило Clang, а разница между дебаг-секциями достигла 2.7 гб.

Важно учесть, что Clang предоставляет простой запуск оптимизаций во время линковки LTO, в то время как GCC имеет серьезные сложности с использованием LTO как на `ld` (ошибка маппинга большого объектного файла), так и на `lld` [9]. При этом GCC не защищен от неопределенного поведения: ярким примером служит бесосновательно долгий процесс компиляции больших файлов, который может затянуться на часы.

Таким образом, хоть существенных изменений в производительности между Clang и GCC нет, применение Clang предлагает определенные преимущества:

- уменьшение размера исполняемого файла;
- обеспечение более стабильного процесса компиляции и линковки;
- возможность применения оптимизаций во время линковки LTO на больших проектах;
- возможность использования статических анализаторов кода;
- возможность дополнения функционала компилятора.

Список литературы

- Holt, Alan, and Chi-Yu Huang. Embedded operating systems. Springer, 2014 – p.151-169.
- Von Hagen, William. The definitive guide to GCC. – 2011 – p. 15-17.
- Lattner, Chris, and Vikram Adve. The llvm compiler framework and infrastructure tutorial. Languages and Compilers for High Performance Computing: 17th International Workshop, LCPC 2004, West Lafayette, IN, USA, September 22-24, 2004, Revised Selected Papers 17. Springer Berlin Heidelberg, 2005. – p. 15.
- Nacke, Kai, and Amy Kwan. Learn LLVM 17: A beginner's guide to learning LLVM compiler tools and core libraries with C++. Packt Publishing Ltd, 2024 – p. 27-32.
- Appel, Andrew W. SSA is functional programming. Acm Sigplan Notices 33.4 (1998): 17-20.
- Glek T., Hubicka J. Optimizing real-world applications with GCC Link Time Optimization. GCC Developers' Summit. – 2010. – p. 25-45.
- Nacke, Kai, and Amy Kwan. Learn LLVM 17: A beginner's guide to learning LLVM compiler tools and core libraries with C++. Packt Publishing Ltd, 2024 – p. 86-88.
- Vichare A. The Conceptual Structure of GCC. Indian Institute of Technology, Bombay. – 2008 – p. 3-11.
- Проблема LTO lld на GitHub [электронный ресурс]. URL: <https://github.com/llvm/llvm-project/issues/41791> (дата обращения: 30.01.2024).

ЭФФЕКТИВНАЯ синхронизация РАБОТЫ НЕСКОЛЬКИХ ПОЛЬЗОВАТЕЛЕЙ В СЛАЙСЕРЕ

*Ковалькова М.Е.², Гавриков А.А.¹,
Бадалян Р.К.², Дуплин Г.В.³*

*Национальный исследовательский университет
«Высшая школа экономики»,*

¹департамент электронной инженерии,

²департамент компьютерной инженерии,

³департамент прикладной математики

МИЭМ НИУ ВШЭ

Аннотация

В данной статье будет рассмотрена проблема синхронизации действий пользователей при работе в слайсере и методы её решения, а также будут перечислены методы практической реализации предложенных вариантов.

Введение

При разработке программного обеспечения, подразумевающего совместную работу над любого рода проектом в режиме реального времени, возникает задача синхронизации действий пользователей. Сложность заключается в том, что в такого рода системах совсем неочевидным является алгоритм действий при одновременной работе нескольких пользователей над одной и той же сущностью, будь то одно и то же место в тексте при работе с текстовыми документами или изменение одинаковых записей в базе данных [1]. Если говорить о процессе работы группы пользователей в интерфейсе слайсера, могут возникнуть проблемы при совместном доступе к настройкам печати, редактированию размеров и положения модели и её поддержек на столе, а также отображении результата слайсинга.

Проблемы совместного доступа

Первый аспект, который необходимо учесть при реализации многопользовательского функционала в слайсере - обеспечение возможности отслеживания изменений, вносимых каждым участником в режиме реального времени. Для этого можно использовать технологию websockets или другие асинхронные механизмы связи между клиентами и сервером, что позволит пользователям обмениваться данными и наблюдать изменения моделей и параметров печати.

Второй аспект, который следует рассмотреть - разделение доступа. Чтобы избежать конфликтов и несанкционированного изменения данных, необходимо предусмотреть механизм разделения доступа, где каждому пользователю должны быть назначены соответствующие разрешения на редактирование определенных элементов проекта. Это может быть реализовано через управление ролями и правами пользователя или системой контроля доступа.

Третьим аспектом является разрешение конфликтов при одновременном редактировании. Когда несколько пользователей вносят изменения в одну и ту же модель или параметры печати одновременно, могут возникать конфликты. Необходимо внедрить ме-

ханизмы синхронизации и управления версиями, что может включать в себя отслеживание изменений, уведомления о конфликтах и возможность отмены и повторения действий.

Четвертый аспект, на который следует обратить внимание - синхронизация изменений между всеми участниками проекта. Требуется эффективный алгоритм слияния изменений и их передачи между пользователями, необходимо предусмотреть механизмы, позволяющие обновлять данные на клиентах при изменениях на сервере и наоборот.

Анализ существующих подходов

Проблема совместного доступа к какому-либо роду объектам возникает у каждого разработчика, перед которым стоит задача реализации многопользовательского интерфейса. Существует два традиционных подхода к решению данной проблемы, OT и CRDT.

OT (Operational Transformation) — это метод, используемый в распределенных системах для обеспечения согласованности данных, которые могут быть изменены в нескольких экземплярах системы одновременно [2]. Операционное преобразование может быть неэффективным в некоторых случаях. Его использование может привести к дополнительным вычислительным нагрузкам и задержкам в системе, особенно при работе с большими объемами данных или при выполнении сложных операций. Это может ограничивать производительность и масштабируемость системы. Также в некоторых случаях операционное преобразование может приводить к потере информации. Если две операции взаимодействуют в таком образе, что невозможно преобразовать одну операцию в другую, возникает конфликт, и система должна принять решение о том, как разрешить этот конфликт [3]. В результате одна из операций может быть отклонена, что может привести к потере данных или искажению их значения.

CRDT — это подход к разработке и представлению данных, который обеспечивает автономное реплицирование и согласованность данных в распределенных системах. Преимущество CRDT состоит в том, что он обеспечивает высокую доступность и отказоустойчивость данных. Каждая реплика может автономно принимать решения и обновлять данные, даже если связь с другими репликами прервана [4]. Когда связь восстанавливается, CRDT автоматически объединяет изменения и обеспечивает согласованность данных без необходимости централизованного согласования. CRDT также обладает хорошей масштабируемостью. Поскольку каждая реплика может выполнять операции над данными локально и самостоятельно, система может легко масштабироваться путем добавления новых реплик без необходимости координации и согласования с существующими репликами [4, 5].

Учитывая специфику алгоритмов OT, подход CRDT менее избыточен с точки зрения спектра решаемых им задач.

Предлагаемый подход

При более детальном рассмотрении процесса многопользовательской работы в интерфейсе слайсера можно выделить четыре аспекта, требующих описания процесса синхронизации. В первом случае одновременно подвергаются изменению разные объекты или разные свойства одного объекта (см. рис.1). Например: первый пользователь изменяет масштаб модели, а второй ее расположение. При этом действия изменения не противоречат друг другу, а поэтому оба изменения будут объединены, для получения итогового состояния объекта.

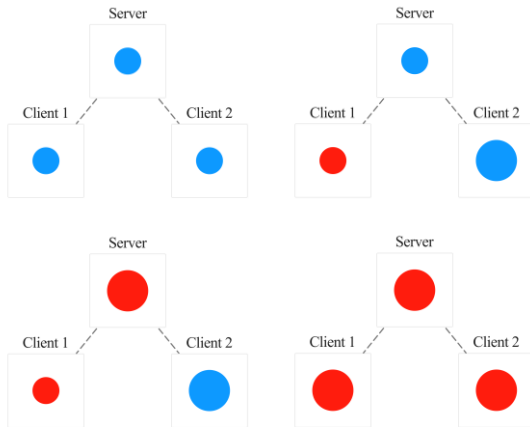


Рис.1. Поэтапное изменение разных свойств объекта

В том случае, когда изменяется одно и то же свойство одного объекта, возникает необходимость разрешения конфликта между двумя состояниями. Самым простым способом будет применять изменения по порядку их поступления (см. рис.2). Тогда итоговое состояние объекта будет задано последним внесенным изменением.

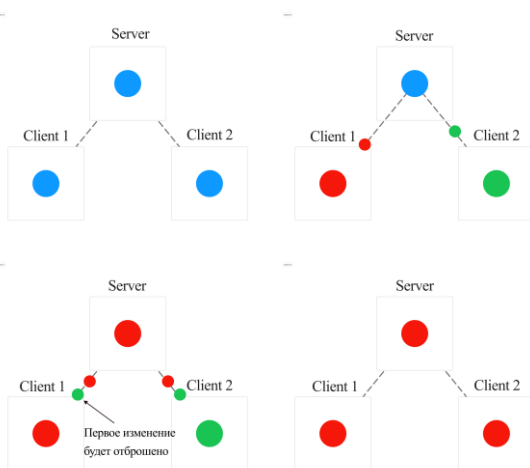


Рис.2. Поэтапное изменение одного и того же свойства объекта

Отдельно стоит упомянуть возможность отмены и повтора действий. Учитывая, что над одним и тем же объектом могут работать несколько пользователей, справедливо будет ограничить возможность отмены действий других членов команды. Соответственно, чтобы решить данную проблему, достаточно для каждой пары пользователь-объект вести отдельный журнал изменений. То есть действие отмены/повтора будет переводить объект в предшествующее/следующее состояние из определенного множества состояний, измененных конкретным пользователем.

При работе в интерфейсе слайсера, пользователь взаимодействует с набором объектов, находящихся в иерархии, состоящих из родительских и дочерних элементов. Такая система предполагает возможность взаимной смены ролей, что в свою очередь создает вероятность возникновения циклов. Например, пользователь может оказаться в ситуации, когда он одновременно отправил серверу неподтвержденное изменение родительского элемента А в разделе В, а также получил изменение от сервера, родительского элемента В в разделе А (см. рис.3). Изменение в будущем будет отклонено сервером, потому что оно формирует цикл, но клиент еще не знает об этом.

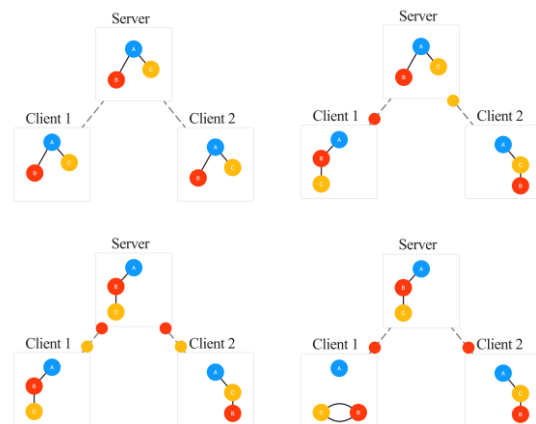


Рис.3. Поэтапное изменение набора объектов, находящихся в иерархии

Заключение

В процессе исследования были рассмотрены проблемы синхронизации действий клиентов многопользовательских приложений и необходимых для их реализации ресурсов. Были конкретизированы основные аспекты, требующие отдельного внимания при разработке программного обеспечения слайсера. На основе рассмотренных подходов к решению обозначенных проблем, для каждого класса задач была предложена соответствующая методика.

Список литературы

1. Imine A. et al. Proving correctness of transformation functions in real-time groupware //ECSCW 2003: Proceedings of the Eighth European Conference on Computer Supported Cooperative Work 14–18 Septem-

ber 2003, Helsinki, Finland. – Dordrecht : Springer Netherlands, 2003. – С. 278.

2. Sun C., Ellis C. Operational transformation in real-time group editors: issues, algorithms, and achievements //Proceedings of the 1998 ACM conference on Computer supported cooperative work. – 1998. – С. 59-68.

3. Preguiça N., Baquero C., Shapiro M. Conflict-free replicated data types (CRDTs) //arXiv preprint arXiv:1805.06358. – 2018.

4. Li D., Li R. An admissibility-based operational transformation framework for collaborative editing systems //Computer Supported Cooperative Work (CSCW). – 2010. – Т. 19. – С. 1-43.

5. Shapiro M. et al. Conflict-free replicated data types //Stabilization, Safety, and Security of Distributed Systems: 13th International Symposium, SSS 2011, Grenoble, France, October 10-12, 2011. Proceedings 13. – Springer Berlin Heidelberg, 2011. – С. 386.

РАЗРАБОТКА ВИЗУАЛИЗАТОРА ДАННЫХ ВЕКТОРНОГО ФАЙЛА G-CODE

Васильев И.А., Фирстова Н.М.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Целью работы является разработка программного компонента для обеспечения визуализации данных файла траектории (G-Code) в заданной экранной области, допуская масштабирование и сдвиг изображения.

Поставлена задача оптимизировать количество отображаемых данных, не обрабатывая те из них, которые не видны в текущем масштабе, для минимизации задержек при перерисовке изображения.

Компонент предназначен для предпросмотра содержимого файлов заданий и визуализации движения инструмента в составе встраиваемой системы управления для промышленного оборудования компании ООО «Бигпринтер Цифровые Инновации».

Введение

В современной промышленности процесс управления оборудованием становится все более автоматизированным и компьютеризированным. Одним из ключевых компонентов этой системы является визуализация данных траектории, представленной в формате G-Code. G-Code – это стандартный формат файлов, используемый для управления числовым программным оборудованием, таким как станки с ЧПУ и 3D-принтеры [1].

Целью работы является разработка программного компонента, способного визуализировать данные из файла G-Code в заданной экранной области, обеспе-

чивая возможность масштабирования и сдвига изображения. Задача не ограничивается простым отображением данных – необходима оптимизация процесса визуализации и минимизирование задержки при перерисовке изображения.

Для достижения этой цели необходимо разработать алгоритм обработки данных, который бы исключал из отображения невидимые в текущем масштабе части траектории. Такой подход позволит значительно улучшить производительность визуализатора, освободив ресурсы для более качественного представления данных.

Разрабатываемый компонент предназначен не только для удобного предпросмотра содержимого файлов заданий, но и для визуализации движения инструмента в составе встраиваемой системы управления промышленным оборудованием. Таким образом, он является незаменимым инструментом для проектировщиков, операторов и технических специалистов, работающих с ЧПУ-станками и аналогичным оборудованием.

Аналоги разрабатываемой системы

На рынке существует несколько аналогичных систем, предназначенных для визуализации данных траектории в формате G-Code. Многие из них не доступны на территории Российской Федерации. Вот несколько примеров аналогов разрабатываемого программного компонента:

Mach3 и Mach4: системы являются популярными среди пользователей станков с ЧПУ. Они обеспечивают визуализацию траектории, но у них есть недостаток – проблема производительности при работе с большими файлами G-Code [2].

LinuxCNC – это свободное программное обеспечение с открытым исходным кодом для управления станками с ЧПУ. Включает в себя инструменты для визуализации траектории [3], но, как и в случае с Mach3/Mach4, возникают проблемы с производительностью при обработке больших файлов.

Несмотря на существующие аналоги, разрабатываемая система предлагает улучшенные возможности визуализации данных G-Code и оптимизированную производительность. Разрабатываемый компонент ориентирован на минимизацию задержек при обработке больших файлов и обеспечение точного и надежного предпросмотра траектории, что делает его удобной для работы с ЧПУ-оборудованием.

Предлагаемое решение

Задачей является разработка визуализатора данных файла G-Code. В распоряжении имеется гетерогенный чип Zynq-7000 (2 ядра ARM на 1 ГГц и FPGA) и Linux – одно ядро, без графического ускорителя. Поставленная задача разделена на несколько подзадач: парсер, визуализация данных в виде рисунка, ограничение ресурсов при выполнении программы, отсечение фигур в Qt, масштабирование.

Для чтения и обработки данных из файла GCode разработан специализированный парсер. Этот компонент отвечает за анализ содержимого файла и из-

влечение необходимой информации о координатах и командах движения инструмента.

Чтобы получить оптимальную визуализацию данных из файла G-Code, в проекте используется класс QPainter из фреймворка Qt [4, 5]. С его помощью можно создать изображение, представляющее собой траекторию движения инструмента. Это позволяет не только эффективно отображать данные, но и применять различные методы масштабирования и сдвига изображения.

Для обеспечения стабильной работы и предотвращения избыточного потребления ресурсов используется функция setrlimit() [6, 7]. Это позволяет установить ограничение на размер оперативной памяти, что в свою очередь помогает предотвратить перегрузку системы.

Методы отсечения фигур в QPainter применены для оптимизации процесса визуализации. Это позволяет исключить из отображения те части траектории, которые не видны в текущем масштабе путем использования методов: setClipRect(), setClipRegion() и setClipPath() [4], в зависимости от сложности формы области отсечения.

Для реализации функционала масштабирования используются возможности класса Qt: QPinchGesture [8]. Этот класс предоставляет информацию о масштабировании, основанную на жестках пользователя, что позволяет удобно управлять масштабом изображения.

Дальнейшее развитие разработанной системы

Предварительные оценки позволяют говорить о том, предлагаемая реализация алгоритмов позволит облегчить работу системы и обеспечить меньшую задержку работы подсистемы визуализации. Тем не менее требуется тестирование разработанного решения с другими функциями (в том числе управление дюзами принтера, перемещением печатающей головки и другими механизмами принтера), которые выполняет процессорная подсистема принтера.

Часть тяжелых задач в существующей реализации отдана FPGA-части Zynq-7000, но она задействована не полностью. Это открывает возможность реализовать часть наиболее медленных функций подсистемы визуализации также на FPGA (например, парсера, как это сделано в работе [9]).

Заключение

В результате разработки программного компонента, создан визуализатор данных файла G-Code, обладающий оптимизированной производительностью. Использование специализированного парсера для чтения и обработки данных G-Code, совместно с классом QPainter из фреймворка Qt для визуализации, позволяет точно и надежно отображать траекторию движения инструмента. Методы оптимизации, такие как ограничение ресурсов, отсечение фигур и масштабирование, обеспечат эффективный способ визуализации изображения.

Список литературы

1. Latif K. et al. A review of G code, STEP, STEP-NC, and open architecture control technologies based embedded CNC systems //The International Journal of Advanced Manufacturing Technology. – 2021. – Vol. 114. – P. 2549-2566.
2. Сравнение Mach 3 и Mach 4 для управления ЧПУ // CNC Maniac URL: <https://cnc-maniac.ru/sravnenie-mach-3-i-mach-4-dlja-upravlenijachpu/> (дата обращения: 09.02.2024).
3. LinuxCNC // LinuxCNC URL: <https://linuxcnc.org/> (дата обращения: 09.02.2024).
4. QPainter Class // Qt documentation URL: <https://doc.qt.io/qt-6/qpainter.html> (дата обращения: 07.02.2024).
5. Eng E. Qt GUI Toolkit: Porting graphics to multiple platforms using a GUI toolkit //Linux Journal. – 1996. – Vol. 1996. – №. 31es. – P. 2-es.
6. Set resource limit for your C++ program // codeforces URL: <https://codeforces.com/blog/entry/82541> (дата обращения: 04.02.2024).
7. FAR S. O. Linux System Calls //Linux Journal. – 2000. – Vol. 75.
8. QPinchGesture Class // Qt documentation URL: <https://doc.qt.io/qt-6/qpinchgesture.html> (дата обращения: 01.02.2024).
9. Балгуданян И.А. и др. Разработка 3D-принтера на основе системы на кристалле с ПЛИС // Системный администратор. – 2020. – № 6(211). – С. 44-49.

Секция «Материалы и технологии»

ТЕХНОЛОГИЯ ЛАЗЕРНОГО СКРАЙБИРОВАНИЯ СОЛНЕЧНОГО ЭЛЕМЕНТА НА ОСНОВЕ ПЕРОВСКИТА

Иштеев Р.Р.^{1,2}, Кондратенко В.С.³
Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент электронной инженерии,

²кафедра информационной безопасности
киберфизических систем

МИЭМ НИУ ВШЭ,

³Институт перспективных технологий и
индустриального программирования
РТУ-МИРЭА

Аннотация

За последнее десятилетие эффективность галогенидных перовскитных солнечных элементов (ПСЭ) быстро возросла до 26,1%. Однако, чтобы масштабировать ПСЭ, используют модульную геометрию, аналогичную солнечным элементам a-Si. В данной работе предложена технология лазерного скрайбирования проводящего слоя $\text{In}_2\text{O}_3:\text{SnO}_2$ и фотоактивного слоя перовскита $\text{Cs}_{0,2}(\text{CH}(\text{NH}_2)_2)_{0,8}\text{PbI}_3$.

Введение

Перовскитные солнечные элементы (ПСЭ) представляют собой структуру, состоящую из тонкоплёночных слоев с p-i-n архитектуры, с поглощающим слоем перовскита CsFAPbI_3 .

Перовскит – новый класс прямозонных полупроводниковых материалов. Имеет структуру ABX_3 , где, изменяя галоген X можно контролировать ширину запрещенной зоны в диапазоне от 1,53 эВ до 3,1 эВ. Элемент A составляет органические или неорганические катионы ($\text{Cs}_{0,2}(\text{CH}(\text{NH}_2)_2)_{0,8}$ или CsFA, B – катион двухвалентного металла (Sn^{+2} , Pb^{+2}) [1]–[6]. Благодаря этим качествам перовскит получил развитие со стороны научного сообщества за последние 24 года в фотоприемниках, приблизившись к эффективности солнечных элементов на основе кремния (~26.1 %). Низкая себестоимость и высокая эффективность перовскитных солнечных элементов дают предпосылки к коммерческому применению перовскитов в качестве возобновляемых источников энергии [7], [8].

При масштабировании лабораторных образцов ($0,15 \text{ см}^2$) до промышленного модуля ($>1 \text{ см}^2$) неизбежно наблюдается рост поверхностного сопротивления электродных слоев. Выходные характеристики напрямую зависят от потерь на сопротивление. Увеличение поверхностного сопротивления приводит к снижению эффективности фотопреобразователя. Паттернирование монолитного модуля позволяет выделить полезную площадь ячеек фотопреобразователя и снизить поверхностное сопротивление для переноса заря-

да через планарные электроды путем изменения соотношения длины и ширины ячейки. Поверхностное сопротивление рассчитывается по формуле (1)[9]–[11]:

$$SR[\Omega/\square] = R[\Omega] \cdot \frac{L[m]}{W[m]} \quad (1)$$

где, $SR[\Omega/\square]$ – поверхностное сопротивление слоя;

$R[\Omega]$ – сопротивление;

$L[m]$ – длина слоя;

$W[m]$ – ширина слоя.

Технология паттернирования подразумевает выделение рабочей области ячеек из общей структуры модуля путем сегментирования составных ячеек. Данную технологию можно разделить на 3 процесса (рис. 1А): P1, P2 и P3. P1 процесс – это удаление нижнего электродного слоя ИТО ($\text{SnO}_2:\text{In}_2\text{O}_3$) толщиной 220 нм. Ключевыми параметрами процесса P1 является отсутствие трещин на поверхности стекла и отсутствие града вдоль кромки, которые могут привести к короткому замыканию нижнего и верхнего электродов. Под градом подразумевается неспарившийся материал вдоль линии скрайба после лазерной обработки, который наплыл поверх основной обрабатываемой пленки. Далее наносится последовательность транспортных и поглощающих слоев толщиной 450 нм и ее скрайбируют (процесс P2), ключевыми параметрами являются целостность предшествующего слоя ИТО и изоляция смежных ячеек для обеспечения последовательного соединения. Процесс P2 обеспечивает контакт между нижним и верхним электродным слоем и обеспечивает последовательное соединение между ячейками. Нанесения верхнего металлического электрода (толщина 100 нм) выполняется методом напыления через трафаретную маску.

В данной работе рассмотрен метод лазерного паттернирования пленок ИТО (процесс P1) и перовскита (процесс P2).

Исследования выполнялись за следующими лазерными установками:

1) Ультрафиолетовый лазер (УФ) NANIO-355: длина волны - 355 нм, максимальная частота импульсов - 100 кГц, мощность излучения - 3 Вт, длительность одиночного импульса - 40 нс.

2) Фемтосекундный лазер ТЕТА-10: максимальная мощность излучения - 10 Вт, максимальная частота импульсов - 200 кГц, длительность импульса - 250 фс, длина волны излучения - 1030 нм.

Ультрафиолетовый лазер имеет меньшую пиковую мощность импульса, чем ИК-лазер, а длина волны находится в ультрафиолетовом диапазоне. Этого достаточно для удаления пленки перовскита минимизируя тепловую передачу к смежным узлам решетки ИТО и предотвращая ее разрушение. Определены параметры (скорость, число проходов и мощ-

ность) работы лазеров промышленного назначения для выполнения процессов P1 и P2.

Эксперимент

Для исследования режимов резки использовались ПСЭ с перовскитом $Cs_{0.2}(CH(NH_2)_2)_{0.8}PbI_3$.

Исследование параметров скрайбирования пленки ИТО (процесс P1) выполнялось фемтосекундным лазером нанесением линий скрайбов вдоль всей подложки с диапазоном скорости 166-500 мм/с с шагом 83 мм/с и при диапазоне мощности от 0.14 до 0.68 Вт с частотой импульсов 100 кГц. Перемещение лазерного излучения осуществлялось системой вращения двух зеркал. Удаление большей части слоя ИТО и сохранение целостности стеклянной подложки было достигнуто при мощности излучения 0.29 Вт.

На основе полученных данных по однопроходному режиму были проведены исследования гибридных режимов резки пленки ИТО (процесс P1). Особенностью гибридного режима является удаление образованного града вдоль кромок линий резки. Гибридный режим подразумевает совмещение трехпроходного режима резки при мощности 0.29 Вт и скорости 417 мм/с и двухпроходного режима при мощности 0.14 Вт и 417 мм/с.

Для определения режимов лазерной резки пленки перовскита рассматривался диапазон мощности 0.3-1 Вт с шагом 0.1 Вт при диапазоне скорости обработки 100-115 мм/с с шагом 1 мм/с (процесс P2). Наилучший результат был достигнут трехпроходным режимом при мощности резки 0.4 Вт и скорости 110 мм/с с приподнятым фокусом на 250 мкм над поверхностью пленки перовскита. Исследование на распределение элементов (PЭ) пленок ИТО и перовскита в месте реза продемонстрировало удаление всего слоя перовскита при анализе на наличие свинца и индия вдоль всей линии скрайба (1).

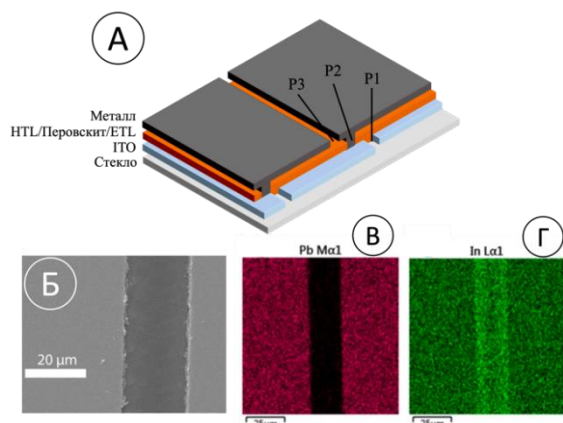


Рис.1. Паттернирование слоя перовскита УФ-лазером мощностью 0,4 Вт, скорость обработки 110 мм/с при поднятии фокуса на 250 мкм трехпроходным режимом: А) Схема паттернирования ПСЭ; Б) СЭМ изображение P2 при фокусировки пучка на поверхности перовскитной пленки; В) Картирование PЭ для In как компонента пленки ИТО; Г) картирование PЭ для Pb как компонент пленки ИТО

Выполнение поставленных задач валидировалось путем внедрения технологии скрайбирования в процесс сборки устройства и сравнении выходных характеристик.

Измерения ВАХ ПСЭ проводились с использованием солнечного имитатора измеритель-источнике Keithley 2400 с разверткой напряжения в диапазоне от 0 до 1,3 В в условиях АМ 1,5G.

Коэффициент полезного действия (КПД) устройств были рассчитаны по формуле 2. Выходные характеристики устройств приведены в таблице 1.

$$\eta = \frac{J_{кз} \cdot U_{xx} \cdot K3}{P_{вх}} \quad (2)$$

где η – КПД, %;

$P_{вх}$ – интенсивность падающего света, мВт/см²;

$J_{кз}$ – плотность тока короткого замыкания, мА/см²;

U_{xx} – напряжение холостого хода, В;

$K3$ – фактор заполнения, %.

Коэффициент заполнения (КЗ) – это параметр, определяющий максимальную мощностью солнечного элемента. КЗ определяется отношением максимальной мощности солнечного элемента к произведению U_{xx} и $J_{кз}$.

Гибридный режим продемонстрировал схожий результат в 17.54% (Таблица 1). В качестве эталона был изготовлен перовскитный солнечный элемент с таким же размером ячейки на подложке 25x25 мм без использования лазерного скрайбирования. Этот солнечный элемент показал ВАХ 18%, что идентично показателю образца сравнения.

Таблица 1. Выходные характеристики перовскитных солнечных элементов с различными методами обработки ИТО-пленки фемтосекундным лазером

| Режимы фрагментации пленки ИТО | U_{xx} (В) | $J_{кз}$ (mA/cm ²) | КЗ (%) | КПД (%) |
|--------------------------------|--------------|--------------------------------|--------|---------|
| Трехпроходный | 1.00 | -26.43 | 67.98 | 18.04 |
| Гибридный | 1.03 | -23.16 | 73.68 | 17.54 |

Наилучший результат паттернирования пленки перовскита (процесс P2) продемонстрировал УФ-лазер при трехпроходном режиме резки мощностью 0.4 Вт и скоростью 110 мм/с, фокусе поднят на 250 мкм от поверхности пленки перовскита.

Наилучший результат резки пленки ИТО (процесс P1) продемонстрировал фемтосекундный лазер. Оптимальным методом обработки пленок ИТО фемтосекундным лазером с длиной волны 1030 нм является гибридный режим. Собранный устройство продемонстрировало КПД 17.54% при КЗ 73.68%. В данном режиме резки пленки ИТО высота града составила 20 и 30 нм по краям линии среза.

Список литературы

1. A. Amat *et al.*, “Cation-induced band-gap tuning in organohalide perovskites: Interplay of spin-orbit coupling and octahedra tilting,” *Nano Letters*, vol. 14, no. 6, pp. 3608–3616, 2014. doi: 10.1021/nl5012992.
2. S. D. Stranks *et al.*, “Electron-hole diffusion lengths exceeding 1 micrometer in an organometal trihalide perovskite absorber,” *Science (80-.)*, vol. 342, no. 6156, pp. 341–344, 2013, doi: 10.1126/science.1243982.
3. I. Artur, K. Constantinova, D. Saranin, and A. Di Carlo, “Investigation of structural and optical properties of MAPbBr₃ monocrystals under fast electrons irradiation,” *J. Mater. Chem. C*, vol. accepted, 2022.
4. N. Li, X. Niu, and Q. C. H. Zhou, “Towards commercialization: the operational stability of perovskite solar cells,” *Roal Soc. Chem.*, pp. 4–10, 2020.
5. Y. Chen *et al.*, “Extended carrier lifetimes and diffusion in hybrid perovskites revealed by Hall effect and photoconductivity measurements,” *Nat. Commun.*, vol. 7, p. 12253, 2016, doi: 10.1038/ncomms12253.
6. L. M. Herz, “Charge-Carrier Mobilities in Metal Halide Perovskites: Fundamental Mechanisms and Limits,” *ACS Energy Lett.*, vol. 2, no. 7, pp. 1539–1548, 2017, doi: 10.1021/acsenerylett.7b00276.
7. “NREL Efficiency Chart,” *Www.Nrel.Gov*. <https://www.nrel.gov/pv/cell-efficiency.html>
8. L. K. Ono and N.-G. Park, “Perovskite Solar Cells - Towards Commercialization,” *ENERGY Focus*, pp. 1749–1751, 2017, doi: 10.1021/acsenerylett.7b00517.
9. L. A. Castriotta *et al.*, “Light-Stable Methylammonium-Free Inverted Flexible Perovskite Solar Modules on PET Exceeding 10.5% on a 15.7 cm² Active Area,” *ACS Appl. Mater. Interfaces*, vol. 13, no. 25, pp. 29576–29584, Jun. 2021, doi: 10.1021/acсами.1c05506.
10. L. Liang *et al.*, “Patterning Technologies for Metal Halide Perovskites: A Review,” *Adv. Mater. Technol.*, vol. 8, no. 3, p. 2200419, Feb. 2023, doi: <https://doi.org/10.1002/admt.202200419>.
11. В. С. Кондратенко, А. Р. Иштеев., Р. Р. Иштеев, и М. К. Константинова, “Исследование лазерного паттернирования компонентов перовскитного фотоприемника,” *Оптические технологии, материалы и системы (Оптотехника - 2021)*, pp. 102–107, 2021.

ОПРЕДЕЛЕНИЕ ДЕФЕКТНОСТИ ПОДЗАТВОРНОГО ДИЭЛЕКТРИКА МДП-СТРУКТУР

*Корнев С.А., Андреев В.В.
МГТУ им. Баумана, Калужский филиал,
факультет «Информатика и управление»
кафедра «Проектирование и технология
производства электронных приборов»*

Аннотация

Описана методика определения параметров надёжности диэлектрика в МДП-структурах, на примере анализа распределения дефектов в диэлектрике

МДП-конденсаторов в зависимости от инжектированного в них заряда.

Введение

Уменьшение проектных норм является глобальным трендом при производстве микросхем, так как это позволяет увеличить степень интеграции и тем самым увеличить их функциональность. Тем не менее, уменьшение проектных норм приводит к увеличению влияния дефектов на параметры полупроводниковых устройств. Поэтому обеспечение и контроль надёжности становятся одной из важнейших задач при производстве микросхем.

Исследованию дефектов подзатворного диэлектрика посвящено большое число публикаций [1-11]. Свойства и качество подзатворного диэлектрика играют первоочередную роль в обеспечении надёжности элементов микросхем, поэтому исследование методов оценки его характеристик и является целью данной работы.

Методы исследований

Для проведения исследования были применены МДП-конденсаторы, изготовленные на основе термического диоксида кремния (SiO₂) на подложках из монокристаллического кремния, легированного фосфором. Пленка диоксида кремния была получена путем термического окисления кремния при наличии кислорода, при температуре от 850 до 1000 °С и добавлением 3% HCl. Толщина полученной пленки диоксида кремния составляла от 30 до 50 нм. Для верхнего электрода площадью от 10⁻⁴ до 10⁻² см² использовалась пленка поликремния (20 Ом/□), толщиной 0,6 мкм с легированием фосфором [1].

Дефекты в диэлектрике можно разделить на внешние и внутренние [7-9].

Внешние дефекты обусловлены процессом производства, в частности, наличием различного рода загрязнений, механическими напряжениями и др. Данный вид дефектов значительно влияет на выход из строя полупроводниковых устройств на ранних этапах эксплуатации или этапах термоциклирования и электротермотренировки, так как они значительно ослабляют диэлектрические свойства [7-9].

Внутренние дефекты, такие как зарядовые дефекты, возникают в диэлектрике в результате приложения электрического поля и протекания тока Фаулера-Нордгейма [11]. При достижении определенного критического уровня внутренних дефектов они вызывают образование проводящего пути в диэлектрике и, следовательно, его пробой [5-6].

Для оценки дефектности МДП-конденсаторов (тестовых структур) применялся метод управляемой токовой нагрузки [1], при котором величина заряда Q_{inj}^{BD} , инжектированная до пробы, измеряется при постоянной плотности тока.

Экспериментальные результаты и обсуждения

На рис.1 представлена относительная гистограмма распределения МДП-конденсаторов по величине Q_{inj}^{BD} .

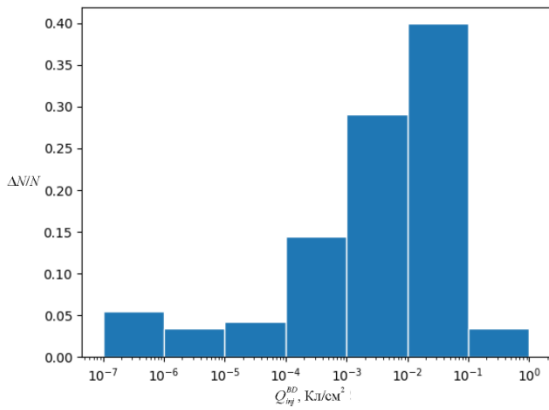


Рис. 1. Относительная гистограмма распределения МДП-конденсаторов по величине Q_{inj}^{BD}

С помощью данных о плотности вероятности полученных с помощью данной гистограммы, был построен график функции распределения образцов по величине Q_{inj}^{BD} .

Для получения аналитического представления функции распределения образцов использовался метод наименьших квадратов, с предположением о том, что искомая функция следует распределению Вейбулла [4].

$$F(Q_{inj}^{BD}) = 1 - \exp\left(-\left(\frac{Q_{inj}^{BD}}{\alpha}\right)^\beta\right), \quad (1)$$

где α – параметр масштаба распределения, β – параметр формы распределения. Далее преобразуем (1) к линейному виду прологарифмировав её.

Для нахождения параметров α и β , примем для координатной плоскости, что $Y = \ln\left(\ln\left(\frac{1}{1-F(Q_{inj}^{BD})}\right)\right)$, $X = \ln Q_{inj}^{BD}$ и $\lambda = -\beta \ln \alpha$, тогда функция Вейбулла преобразуется к линейному виду.

Данное линейное уравнение используется для аппроксимации функции распределения тестовых образцов [3]. Функцию распределения по Q_{inj}^{BD} можно рассмотреть как совокупность функций распределения, связанных с внешними ($F(Q_{inj}^{BD})_{External}$) и внутренними ($F(Q_{inj}^{BD})_{Internal}$) дефектами [2].

Обозначим точку пересечения $F(Q_{inj}^{BD})_{External}$ и $F(Q_{inj}^{BD})_{Internal}$, как $F_{FailureExternal}$ (F). Данная точка определяет процент дефектных структур, которые обусловлены внешними дефектами. С помощью неё можно рассчитать процент выхода годных Y , вычтя значение F из единицы.

При условии, что внешние дефекты распределены по пластине по закону распределения Пуассона [10], можно вычислить их плотность D_s :

$$D_s = \frac{-\ln(1-F)}{S}, \quad (2)$$

где S – площадь тестовой структуры.

Найденные параметры (F , Y и D_s) могут использоваться для оценки качества технологического процесса.

Заключение

Данная методика позволяет определить основные характеристики с помощью, которых можно оценить качество диэлектрика. Также она позволяет контролировать процесс производства интегральных схем, выявлять и корректировать отклонения в нём.

Список литературы

1. Andreev D.V., Maslovsky V.M., Andreev V.V., Stolyarov A.A. Modified Ramped Current Stress Technique for Monitoring Thin Dielectrics Reliability and Charge Degradation // Phys. Status Solidi A. 2022. Vol. 219. Is. 9. P. 2100400(1-5). <https://doi.org/10.1002/pssa.202100400>
2. Сивченко А.С. Методика определения дефектности подзатворного диэлектрика с использованием ускоренных испытаний тестовых структур // Изв. вузов. Электроника. 2015. №3. С.304-312.
3. Martin A., Vollertsen R.-P., Mitchell A., Traving M., Beckmeier D., Nielen H. Fast wafer level reliability monitoring as a tool to achieve automotive quality for a wafer process. Microelectronics Reliability, 2016, v. 64. pp. 2-12.
4. Strong A., Wu E., Vollertsen R., Sune J., Rosa G., Rauch S., Sullivan T. Reliability Wearout Mechanisms in Advanced CMOS Technologies. IEEE Press Series on Microelectronic Systems. Wiley, 2009, 624 p.
5. Kim A., Wu E., Li B., Linder B. Transformation of Ramped Current Stress V_{BD} to Constant Voltage Stress TDDB T_{BD} . IEEE International Reliability Physics Symposium (IRPS), Monterey, CA, USA, 2019, pp. 1-5.
6. JEDEC Standard 122E, Failure Mechanisms and Models for Semiconductor Devices. 2009.
7. Lombardo S., Stathis J.H., Linder P., Pey K.L., Palumbo F., Tung C.H. Dielectric breakdown mechanisms in gate oxides. J. Appl. Phys., 2005. v. 98. pp. 121301 (1-36).
8. Wu E.Y. Facts and Myths of Dielectric Breakdown Processes—Part I: Statistics, Experimental, and Physical Acceleration Models. IEEE Trans. Electron Devices, 2019, v. 66, pp. 4523-4534.
9. Palumbo F., We C., Lombardo S., Pazos S., Aguirre F., Eizenberg M., Hui F., Lanza M. A Review on Dielectric Breakdown in Thin Dielectrics: Silicon Dioxide, High-k, and Layered Dielectrics. Adv. Funct. Mater., 2019, v. 1900657, pp. 1900657(1-26).
10. McPherson J.W., Khamankar R.B., Shanware A. Complementary model for intrinsic time-dependent dielectric breakdown in SiO2 dielectrics // J. Appl. Phys. 2000. Vol.88. P.5351. <https://doi.org/10.1063/1.1318369>
11. Андреев Д.В., Бондаренко Г.Г., Андреев В.В., Столяров А.А. Повышение зарядовой стабильности пленок подзатворного диэлектрика МДП-структур методом их легирования фосфором // Перспективные материалы. 2020. № 7. С.68-74. DOI: 10.30791/1028-978X-2020-7-68-74

ИССЛЕДОВАНИЕ НАПРАВЛЕННЫХ ОТВЕТВИТЕЛЕЙ НА БАЗЕ ИНТЕГРАЛЬНОЙ ОПТИКИ НА НИТРИД КРЕМНИЕВОЙ ПЛАТФОРМЕ

Венедиктов И.О.

*Национальный исследовательский университет
«Высшая школа экономики»,
базовая кафедра квантовой оптики
и телекоммуникаций ЗАО «Сконтел»
МИЭМ НИУ ВШЭ*

Аннотация

В работе исследованы направленные ответвители, изготовленные на базе интегральной оптики и изготовленные на подложке из нитрида кремния. Длина участка взаимодействия варьировалась от 16 мкм до 100 мкм. Предметом исследования являлся коэффициент деления направленных ответвителей и влияние на него от глубины травления (110 нм и 131 нм) при пропускании света с длиной волны от 900 до 945 нм.

Введение

Интегральная оптика и нанофотоника позволяет создавать устройства, которые применяются в квантовых вычислениях, квантовой и классической коммуникации. Данная реализация оптических устройств имеет ряд преимуществ по сравнению с волоконной и «объемной» оптикой: малый размер устройств уменьшает температурные и механические флуктуации, приводящие к искажению фазы излучения в устройстве, а также позволяет легко масштабировать устройства.

Элементная база интегральной оптики включает в себя множество частей таких как устройства ввода излучения, фазовые модуляторы, делители мощности. Одной из реализаций делителей оптической мощности являются направленные ответвители, которые представляют из себя 2 волновода, сведенные на расстояние порядка длины волны излучения. Перетекание мощности происходит за счет «эванесцентного поля», которое распространяется вне волновода. Коэффициент деления определяется геометрией ответвителей и материалом волноводов. Направленные ответвители используются для построения интерферометров [1] амплитудных модуляторов, создания квантовых вентилях в фотонных квантовых компьютерах [2], поэтому для проектирования направленных ответвителей с определенным коэффициентом деления важно знать, как этот коэффициент деления будет зависеть геометрии ответвителя и от длины волны излучения. Выбранный для исследования диапазон длин волн связан со спектром испускания квантовых точек GaAs (910-925 нм), которые обладают высокой степенью неразличимости фотонов [3]. Поэтому квантовые точки GaAs являются превосходными кандидатами в качестве источников одиночных фотонов в интегрально-оптических реализациях квантового компьютера, где первым шагом на пути к их реализации, является объединение квантовых точек с направленным ответвителем на одной интегрально-оптической платформе.

Описание устройства

Изучаемые устройства были изготовлены на основе платформы нитрида кремния (Si_3N_4). Толщина кремниевой подложки составляла 525 мкм, толщина слоя SiO_2 – 2,5 мкм, а толщина волноводного слоя Si_3N_4 – 220 нм (глубина травления $h_e = 110$ нм и 131 нм для 2 образцов). Типичное устройство состояло из двух волноводов шириной 700 нм с одной входной и тремя выходными дифракционными фокусирующими решетками для ввода и вывода излучения. Параметры дифракционной решетки были оптимизированы на длину волны 925 нм. Волноводы имели участок (L_i на рис.1), на котором они были сведены на расстояние $h_g = 0,5$ мкм (зазор), где за счет «эванесцентного» поля происходило перетекание оптической мощности из одного волновода в другой (рис.1). Перетекание мощности можно также объяснить интерференцией четной и нечетной мод излучения в ответвителях (рис.2). Так как эти моды имеют разный эффективный показатель преломления, то распространяются по волноводу с разной скоростью и получают разный набег фазы. Меняя длину взаимодействия L_i можно менять набег фазы между модами. Тогда конструктивная или деструктивная интерференция двух этих мод соответствует тому, что все излучение находится в левом или правом волноводу. Для получения данных о длине каплинга был изготовлено массива устройств с глубиной травления $h_e = 110$ нм и 131 нм, где длина L_i варьировалась от 16 мкм до 100 мкм с шагом в 6 мкм.

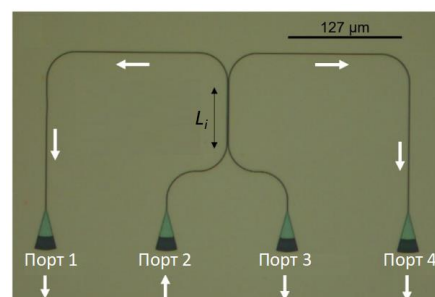


Рис.1. Микрофотография направленного ответвителя

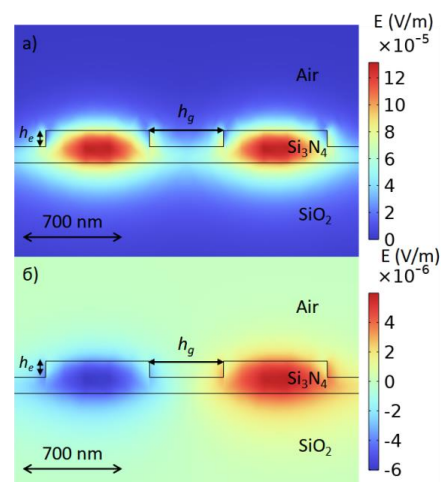


Рис.2. Распределение x -проекции электрического поля для а) четной моды; б) нечетной моды

Методика измерений и результаты

При проведении измерений чип располагался столике с термостабилизацией при помощи элементов Петелье. Также столик был оборудован пьезоподвижками, которые могли перемещать столик вдоль 3 перпендикулярных осей и осуществлять его вращение по оси нормальной к плоскости столика. В качестве источника излучения выступал суперконтинуум лазер Leukos ROCK. Для ввода излучения использовался массив волокон, расположенный над чипом под углом 8 градусов к нормали. Оптическое излучение заводилось в порт 2 (рис.1). С остальных портов при помощи спектрометра Ocean Optics (Maya Pro) измерялся спектр выходной мощности. Порты 1 и 4 были использованы для вычисления коэффициента деления. Порт 3 использовался для оценки отражений, которые составили порядка 3-4% от мощности в основных портах 1 и 4. При вычислении коэффициента деления отражения не учитывались ввиду своей малости, а также поскольку в реальном делителе деление определяется лишь портами 1 и 4. Для каждой длины волны был вычислен коэффициент деления по формуле (1):

$$k = P_1 / (P_1 + P_4) \quad (1)$$

где P_1 и P_4 – оптическая мощность в портах 1 и 4, соответственно. Зависимости коэффициентов деления от длины взаимодействия (рис.5) могут быть аппроксимированы формулой (2):

$$k(L) = 0.5(1 + \cos(\pi L_i / L_C)) \quad (2)$$

где L_i – длина взаимодействия волноводов, а L_C – длина участка взаимодействия, на которой происходит полное перетекание оптической мощности из одного волновода в другой (длина каплинга). Косинусоидальная зависимость легко объясняется через интерференцию четной и нечетной мод в направленном ответвителе (рис.2). Тогда длина каплинга имеет физический смысл длины, при которой разность фаз между четной и нечетной модами составляет π . Анализ экспериментально полученных данных (рис.3 и рис.5) показывает, что длина каплинга больше для структур с большей глубиной травления. Это можно объяснить тем, что с увеличением слоя Si_3N_4 между волноводами уменьшается размер области (воздуха) с показателем отличным от Si_3N_4 , что упрощает перетекание мощности. Также видно, что длина каплинга линейно убывает с длиной волны излучения (рис.4 и рис.6), что связано с уменьшением размера оптической моды, приводящей к ослаблению взаимодействия между волноводами. Стоит отметить, что зависимость длины каплинга от длины волны излучения должна быть экспоненциальной, но поскольку длина волны меняется в малом диапазоне, можно считать данную зависимость линейной в первом приближении.

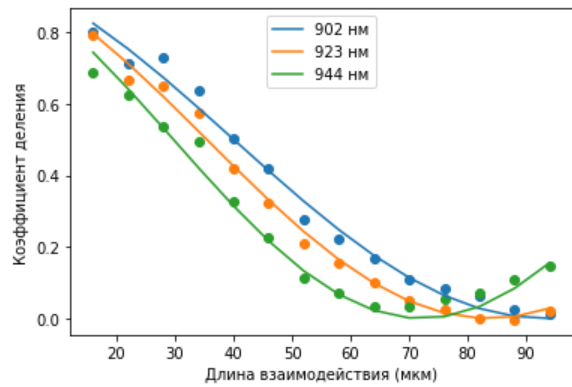


Рис.3. Зависимость коэффициента деления от длины взаимодействия при $h_e = 131$ нм

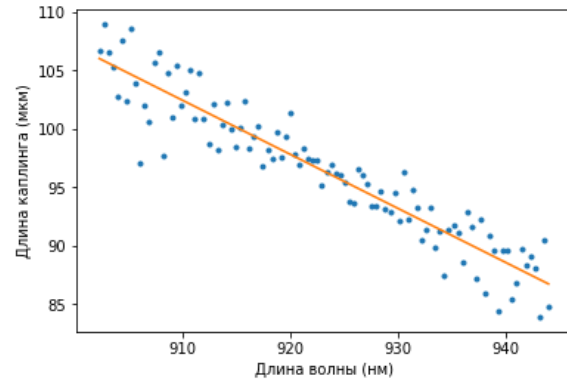


Рис.4. Зависимость длины каплинга от длины волны при $h_e = 131$ нм

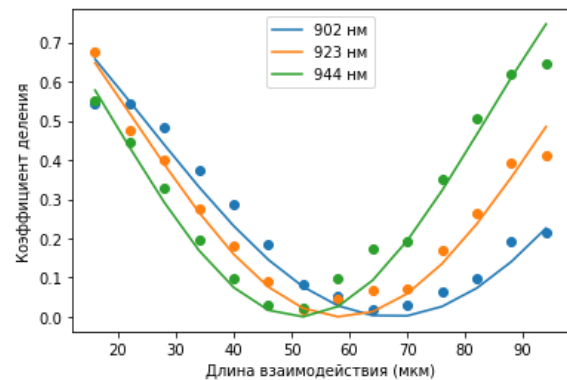


Рис.5. Зависимость коэффициента деления от длины взаимодействия при $h_e = 110$ нм

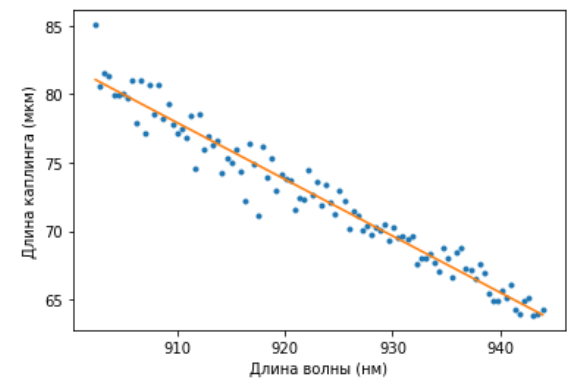


Рис.6. Зависимость длины каплинга от длины волны при $h_e = 110$ нм

Заключение

В данной работе были проведены измерения деления направленных ответвителей, реализованных на базе интегральной оптики на подложке из нитрида кремния. Были получены значения длин взаимодействия 47.76 мкм и 35.85 мкм при глубинах травления 131 и 110 нм соответственно для достижения коэффициента деления 50:50 на длине волны 925 нм. Полученные данные могут быть использованы для изготовления устройств интегральной оптики с направленными ответвителями для преобразования фотонов в оптическом квантовом компьютере с источниками излучения в виде квантовых точек GaAs.

Список литературы

1. A. Prokhodtsov, V. Kovalyuk, P. An, A. Golikov, R. Shakhovoy, V. Sharoglazova, A. Udaltsov, Y. Kurochkin, G. Goltsman. Silicon nitride Mach-Zehnder interferometer for on-chip quantum random number generation. // Journal of Physics: Conference Series – 2020. – Vol. 1695 – 012118.
2. Qiang, X., Zhou, X., Wang, J. et al. Large-scale silicon quantum photonics implementing arbitrary two-qubit processing. // Nature Photonics – 2018 – Vol. 12 – 534–539.
3. Schwartz M. et al. Fully on-chip single-photon Hanbury-Brown and Twiss experiment on a monolithic semiconductor–superconductor platform // Nano Letters. – 2018. – Т. 18. – №. 11. – С. 6892–6897.

ВЛИЯНИЕ ПУЧКОВО-ПЛАЗМЕННОГО ОБЛУЧЕНИЯ НА ПОВЕРХНОСТЬ ВАНАДИЯ И ВЫБРОС ЧАСТИЦ В УСТАНОВКЕ ПЛАЗМЕННЫЙ ФОКУС

*Епифанов Н.А., Дёмин А.С., Морозов Е.В.
Федеральное государственное
бюджетное учреждение науки
Институт металлургии и материаловедения
им. А.А. Байкова Российской академии наук*

Аннотация

Исследована повреждаемость поверхности ванадия и выброс с нее микро-капель в вакуумную камеру установки Плазменный фокус (ПФ) «Вихрь» при воздействии импульсных потоков ионов и плазмы с плотностью мощности излучения $\sim 10^6 - 10^8$ Вт/см² с использованием в качестве рабочего газа дейтерия или гелия. Показано, что в результате импульсного нагрева на поверхности ванадия образуется жидкая пленка – тонкий слой расплава. На поверхности этого расплава возникает неустойчивость Кельвина-Гельмгольца, что приводит к образованию волн в расплавленном слое. Воздействие потока плазмы и мощных акустических импульсов, возникающих при импульсном облучении образца, приводит к срыву микро-капель жидкого ванадия и перенос их на поверхность медного экрана-коллектора. Отмечено возможное влияние подобных процессов на работу установок управляемого термоядерного синтеза.

Введение

Известно, что при воздействии плазмы на конструкционные материалы в вакуумной камере установок термоядерного синтеза (УТС) типа «токамак» образуется значительное количество пылевых частиц, которые могут оказывать негативное влияние на эксплуатацию реактора [1]. В настоящее время полагают, что существуют некоторые пороговые энергии, воздействующие на стенку вакуумной камеры УТС, при которых может расплавиться и испаряться тонкий слой металла с последующей конденсацией в виде пылевых частиц [2]. Ряд механизмов образования пылевых частиц при плазменно-тепловых воздействиях на вольфрам (охрупчивание и механическое разрушение поверхности, капельная эрозия, разрушение поверхностных слоев под действием импульсных плазменных нагрузок, ожидаемых в ИТЭР и др.) рассмотрены в обзоре [3]. В работах [4, 5] рассмотрено влияние геометрии поверхности на характеристики эрозии и выброса микрокапель из вольфрамовых мишеней в условиях воздействия многократных плазменных импульсов; оценена роль гидродинамической неустойчивости поверхностной плёнки жидкой фазы в процессе эрозии.

Целью данной работы было изучение выброса микрочастиц с поверхности технического ванадия марки ВНМ-1 при ее повреждении в процессе воздействия импульсных потоков ионов дейтерия и дейтериевой плазмы с плотностью мощности в интервале $\sim 10^6 - 10^8$ Вт/см² в установке ПФ. Для сравнения проведены также эксперименты по облучению V потоками ионов и плазмы с использованием в качестве рабочего газа гелия.

Методика эксперимента

Облучение образцов ванадия марки ВНМ-1 потоками ионов и плазмы выполняли в установке Плазменный фокус «Вихрь» (ИМЕТ РАН) [6]. Установка ПФ имела медные цилиндрические аксиальные электроды - анод диаметром $\varnothing = 30$ мм, катод $\varnothing = 50$ мм. Энергетический вклад в электрический разряд составлял ~ 3 кДж. Длительность суммарного импульсного воздействия ионно-плазменного потока $\tau \sim 100$ нс. Рабочим газом были дейтерий или гелий при давлении в камере ~ 3 Торр. Плотность мощности q потока ионов и плазмы регулировали в пределах от $\sim 10^6$ до 10^8 Вт/см² путем изменения расстояния L от анода установки до образца и напряжения на конденсаторах U . В качестве базового расстояния брали $L = 8$ см, значение U варьировали в пределах 9 – 10 кВ. Образцы облучали 5 импульсами ионов и плазмы через интервал времени между импульсами ~ 2 мин, в течение которого образцы успевали остывать до исходной температуры. Начальная температура образцов была ~ 300 К.

Образцы ванадия марки ВНМ-1 имели размеры $\sim 1,0 \times 17 \times 17$ мм. Перед облучением образцы шлифовали, полировали пастой ГОИ и промывали этанолом и дистиллированной водой. После подготовки образцов на поверхности ванадия на воздухе образовывалась тонкая плёнка оксида V_2O_5 . Для контроля выброса частиц V при воздействии потока ионов и

плазмы образцы закрепляли в специальном держателе с кольцевым экраном-коллектором (рис.1).

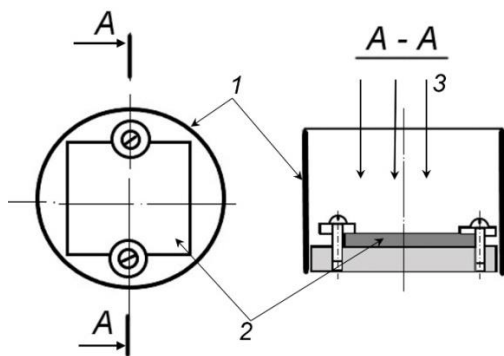


Рис.1. Приспособление для облучения образцов ванадия потоками ионов и плазмы:
1 – цилиндрический экран-коллектор из медной фольги толщиной 0,2 мм, 2 – образец ванадия, 3 – направление падающего пучково-плазменного потока

Результаты экспериментов

Распределение температуры в поверхностном слое

Температура поверхностного слоя образцов оценивалась численным методом [7]. Расчетные температуры нагрева тонких слоев V при плотности мощности дейтериевой плазмы $q = 10^7$ и 5×10^6 Вт/см² приведены на рис. 2 и 3. Численные оценки показали, что при $q = 5 \times 10^6$ Вт/см² температура на облученной поверхности ванадия при каждом импульсном воздействии будет близка к его температуре плавления (~ 1900°C), а при $q = 10^7$ Вт/см² должна существенно (~ на 1500°C) превышать ее.

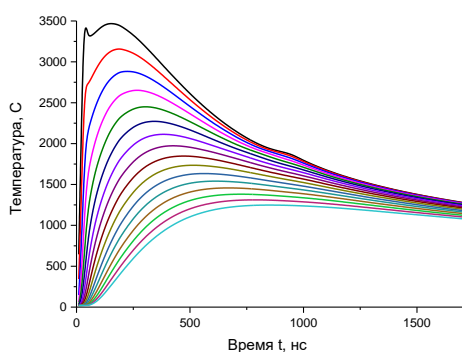


Рис.2. Изменение распределения температуры поверхностного слоя ванадия при импульсном нагреве под действием потоков ионов дейтерия и дейтериевой плазмы с плотностью мощности 10^7 Вт/см². Верхняя кривая описывает изменение температуры на глубине 0,1 мкм от поверхности. Каждая последующая кривая соответствует более глубокому слою с шагом 0,2 мкм

Анализ показал, что при использовании в качестве рабочего газа гелия кривые распределения температуры в поверхностном слое ванадия при тех же значениях плотности мощности практически не

отличаются от кривых для дейтерия, изображенных на рис.2 и 3.

Были также выполнены численные расчеты оценки расстояний образца-мишени от анода, необходимых для получения требуемых значений плотности мощности излучений (в интервале $\sim 10^6$ – 10^8 Вт/см²) при фиксированном напряжении на конденсаторе U и заданной геометрии ионного и плазменного потоков в камере установки ПФ «Вихрь».

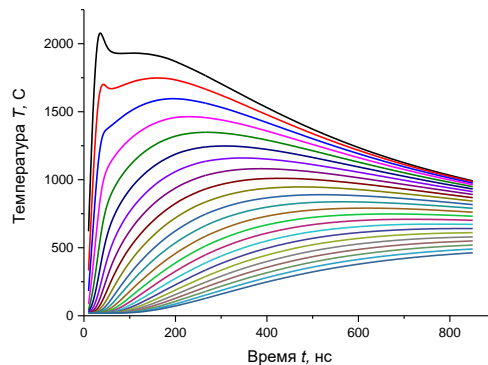


Рис.3. Изменение температуры в поверхностном слое ванадия при импульсном нагреве под действием потоков ионов дейтерия и дейтериевой плазмы с плотностью мощности 5×10^6 Вт/см². Верхняя кривая описывает изменение температуры на поверхности образца

Повреждаемость поверхности ванадия

На рис. 4 показан образец ванадия после облучения ионами дейтерия и дейтериевой плазмой при $q \sim 5 \times 10^6$ Вт/см². Визуально видимых областей оплавления поверхности не наблюдается. На отдельных участках поверхности (рис.4 б, в) видно отслоение пленки основного оксида ванадия V_2O_5 от облученной поверхности.

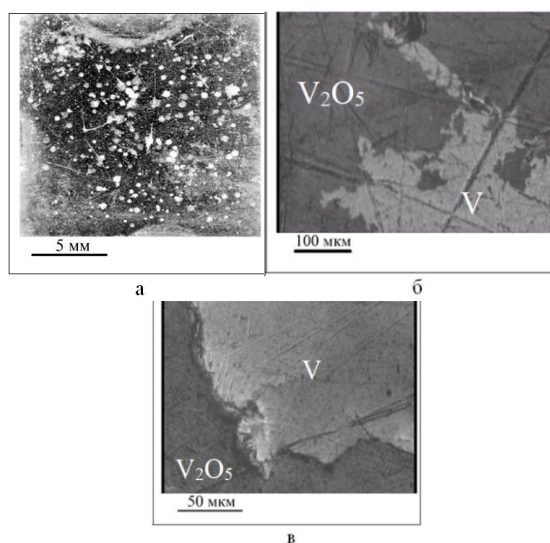


Рис.4. Поверхность ванадия после воздействия дейтериевой плазмы при $q \sim 5 \times 10^6$ Вт/см²: а - общий вид зоны облучения; б, в – микрофотографии облученной поверхности при различном увеличении

При пучково-плазменном воздействии с максимальной плотностью мощности $q \sim 10^8$ Вт/см² практически во всей зоне облученного образца наблюдаются следы интенсивного плавления поверхностного слоя ванадия (рис.5а): на поверхности затвердевшего расплава присутствуют характерные волны, микротрещины и полосы скольжения (рис. 5б, в).

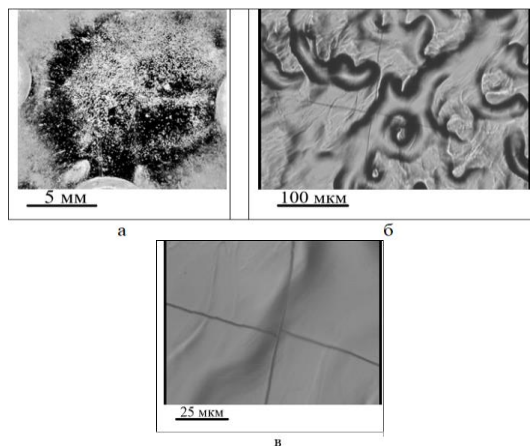


Рис.5. Поверхность ванадия после воздействия дейтериевой плазмы при $q \sim 10^8$ Вт/см²: а – общий вид зоны облучения; б, в – микрофотографии участков волнообразного рельефа при различном увеличении с микротрещинами и следами пластической деформации

Выброс частиц ванадия

Рентгеноспектральный анализ элементного состава медного экрана показал, что при облучении ванадия с использованием обоих рабочих газов в режиме с плотностью мощности $q \sim 10^8$ Вт/см², на внутренней поверхности медного экрана появляются капли ванадия примерно микронного размера. На рис.6а показан участок поверхности медного экрана с осажденными микрокаплями ванадия после его облучения потоками ионов гелия и гелиевой плазмы. Элементный состав капли, обозначенной на снимке «Т2», соответствует ванадию марки ВНМ-1, а состав поверхности в прилегающей к капле области – содержанию элементов в технической меди с сопутствующими примесями (рис.6б).

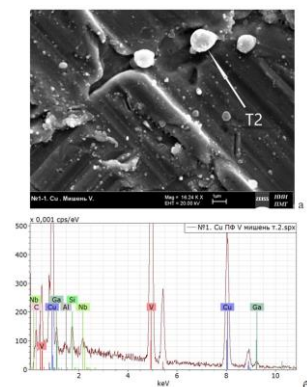


Рис.6. Поверхность медного экрана (а) после 5 импульсных воздействий потоков ионов гелия и гелие-

вой плазмы с плотностью мощности $q \sim 10^8$ Вт/см² на образце ванадия в установке ПФ «Вихрь»

Обсуждение результатов

Повреждение плёнки основного оксида ванадия V₂O₅, наблюдаемое на поверхности ванадия даже при его облучении с минимальным значением плотности мощности q , показало, что порог плавления оксидной пленки лежит вблизи величины $q \sim 5 \times 10^6$ Вт/см². Порог плавления ванадия лежит при более высокой плотности мощности радиационных потоков примерно на порядок величины. Эти оценки подтверждаются результатами численных расчетов распределения температуры в поверхностном слое материала мишени (рис. 2 и 3).

Характер волн, которые образовались на поверхности расплавленного слоя ванадия после его импульсного облучения в установке ПФ, показывает, что их формирование связано с действием горизонтальной составляющей V_x потока плазмы, движущейся параллельно поверхности жидкости. Движение потока плазмы вдоль поверхности расплава приводит к появлению неустойчивости Кельвина-Гельмгольца и образованию волн, имеющих тенденцию к росту амплитуды в области максимальной горизонтальной скорости потока плазмы. Интенсивное динамическое давление потока плазмы на поверхностные волны в сочетании с возможностью при $q \geq 10^8$ Вт/см² ударно-волнового воздействия импульсных потоков энергии на облучаемую мишень способствует срыву отдельных капель с верхней части волн и выбросу этих капель на медный экран-коллектор. Оценка показывает, что роль ударно-волнового воздействия импульсного облучения может сказываться как при ударном давлении плазмы на поверхность расплава, так и при выходе на облучаемую поверхность волны разгрузки, отражённой от задней поверхности образца.

Список литературы

1. Мартыненко Ю.В., Нагель М.Ю. Образование пыли в токамаке. Вопросы атомной науки и техники. Сер. Термоядерный синтез, 2009, вып. 3, с. 43—48.
2. Krasheninnikov S. I., Smirnov R. D., and Rudakov D. L. Dust in magnetic fusion devices. [Plasma Physics and Controlled Fusion](#), 2011, Volume 53, Number 8 (083001), 54 pp. DOI 10.1088/0741-3335/53/8/083001.
3. Будаев В.П. Результаты испытаний вольфрамовых мишеней дивертора при мощных плазменно-тепловых нагрузках, ожидаемых в ИТЭР и токамаках реакторного масштаба (Обзор). Вопросы атомной науки и техники. Сер. Термоядерный синтез, 2015, т. 38, вып. 4, с. 5—33.
4. Мирошник М.О., Геращенко С.С., Махлай В.А., Гаркуша И.Е., Аксенов Н.Н., Бырка О.В., Чеботарев В.В., Кулик Н.В., Лебедев С.И., Шевчук П.Б., Стальцов В.В. Характеристика макроскопической эрозии поверхности зубчатых и плоских мишеней в условиях переходных плазменных нагрузок в ИТЕР. Вопросы Атомной Науки и Техники. Серия: Физика плазмы. 2019. №1(119), с. 70 – 73. DOI: 10.46813/2019-119-070

5. Makhlai V. A., Herashchenko S. S., Aksenov N. N., Byrka O. V., Garkusha I. E., Wirtz M. and Spilker B. Damaging of inclined/misaligned castellated tungsten surfaces exposed to a large number of repetitive QSPA plasma loads. Phys. Scr. T171 (2020) 014047 (9pp). <https://doi.org/10.1088/1402-4896/ab4e52>

6. Грибков В.А., Боровицкая И.В., Демин А. С. и др. Установка “Вихрь” типа Плазменный фокус для диагностики радиационно-термической стойкости материалов перспективных для термоядерной энергетики и аэрокосмической техники. Приборы и техника эксперимента, 2020, № 1, с. 75 – 83.

7. Масляев С.А. Тепловые эффекты импульсного облучения материалов в установке Плазменный Фокус. Перспективные материалы, 2007, № 5, с. 47 – 55.

МОДИФИКАЦИЯ ПОВЕРХНОСТИ ПОЛИМЕРОВ ПОСРЕДСТВОМ ВОЗДЕЙСТВИЯ НА НИХ ВАКУУМНОГО УЛЬТРАФИОЛЕТА

*Ручий А.Ю., Домарев С.Н.,
Рыжих Д.А., Пушкарская Д.В.*

*Белгородский государственный технологический университет им. В.Г. Шухова,
кафедра теоретической и прикладной химии
Химико-технологический институт*

Аннотация

В работе рассматривается влияние вакуумного ультрафиолетового облучения на полиимидную, полиэтилентерефталатовую и фторопластовую плёнки. Изучены массовые потери, электрические параметры и изменение свойств смачиваемости исследуемых материалов.

Введение

В последние годы исследования в области вакуумного ультрафиолета и его воздействия на полимеры привлекли все большее внимание научного сообщества. Вакуумный ультрафиолет (ВУФ) – это область электромагнитного спектра с длиной волны от 100 до 200 нм [1]. Исследователи обнаружили различное воздействие данного излучения на полимерные материалы [2-4].

На полимерный композит на основе полистирола вакуумный ультрафиолет повлиял следующим образом: уменьшилась шероховатость поверхностного слоя материала [5].

В статье [6] изучали воздействие импульсного УФ-облучения на полиэтилен низкой плотности. Исследователи пришли к выводу, что в результате облучения происходят процессы обрыва, сшивания цепей, а также процессы окисления связей поверхностного слоя полиэтилена.

Композит на основе ударопрочного полистирола под воздействием ВУФ-облучения претерпел следующие изменения: изменилась морфология поверхностного слоя материала. Также в исследовании [7] выяснили, что в результате облучения происходит

разрушение верхнего слоя полимера и удаление с поверхности образующихся продуктов.

В статье [8] рассматривается воздействие ВУФ-облучения на полимерный композит на основе полиалканимида. При облучении полиалканимид претерпел массовые потери в размере 0,56 %, а композиты на его основе теряли 0,4 - 0,5 % массы.

Цель исследования данной работы заключается в изучении воздействия вакуумного ультрафиолета (ВУФ) на полимерные плёнки. Работа направлена на оценку изменений, происходящих в полимерных плёнках под воздействием ВУФ, а также на выявление возможных эффектов и свойств, которые могут возникать в результате облучения. В первом разделе описывается метод облучения полимеров в вакуумной ультрафиолетовой установке, а также условия облучения. Во втором разделе представлены результаты проделанной работы.

Методы и материалы

Для исследования воздействия ВУФ на полимерные плёнки были выбраны следующие материалы: полиимидная плёнка (ПИ), полиэтилентерефталатовая плёнка (ПЭТ), фторопластовая плёнка (ФП). Все образцы имеют толщину 20 мкм.

Облучение плёнок производилось в вакуумной ультрафиолетовой установке, представленной на рис. 1.

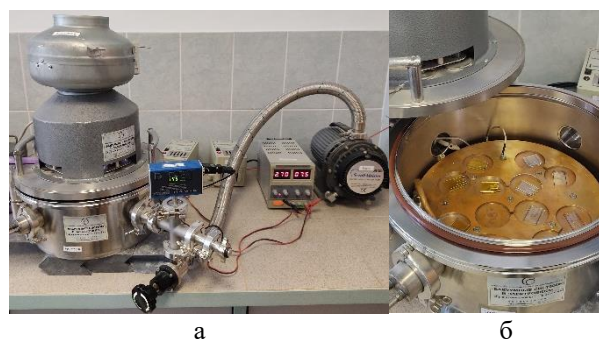


Рис. 1. Внешний вид вакуумной ультрафиолетовой установки: а – установка, б – внутренняя камера

Установка представляет собой сборную конструкцию, состоящую из внутренней камеры (рис. 1б), форвакуумного насоса, отсека с 12 ультрафиолетовыми лампами и лабораторного блока питания. Исследуемые образцы нарезались 25 на 25 мм и размещались на предметном стекле. Затем они закреплялись при помощи нержавеющей проволоочного держателя и помещались в установку. При достижении вакуума в 160 мТорр включались ультрафиолетовые лампы с током 2,70 А и напряжением 7,60 В. Облучение происходило в течение 60 ч монохроматическим светом с длиной волны 90 - 115 нм. Температура внутри камеры составляла 25 °С. Давление в камере контролировали при помощи модульного датчика вакуума GRANVILLE-PHILLIPS Mini-Convectron серии 275. Через каждые 6 часов облучения измерялась масса исследуемых образцов.

Массу образцов измеряли при помощи лабораторных весов CAS CAUW 220D с дискретностью отчета 0,1 мг.

Электрическую ёмкость плёнок измеряли при помощи мультиметра Elitech MM 300 с площадью контакта 4 см², представленного на рис. 2.

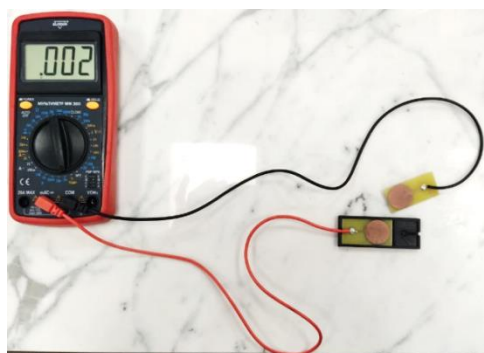


Рис.2. Внешний вид мультиметра

Для анализа изображений капли дистиллированной воды на исследуемых образцах использовалась программа ImageJ с плагином Contact Angle. Полученные результаты обрабатывались в программе Adobe Illustrator.

Основная часть

Массовые потери исследуемых образцов представлены на рис. 3. Согласно рисунку, полиимид теряет основную массу при 48 часах облучения, при этом максимальные массовые потери составили 0,9 %. Полиэтилентерефталат также теряет основную массу при 48 часах облучения, однако, максимальные массовые потери составили 0,83 %. Фторопластовая плёнка теряет основную массу при 54 часах облучения, при этом максимальные массовые потери составили 0,71 %. По результатам анализа можно сказать, что наименьшие массовые потери наблюдаются у фторопластовой плёнки.

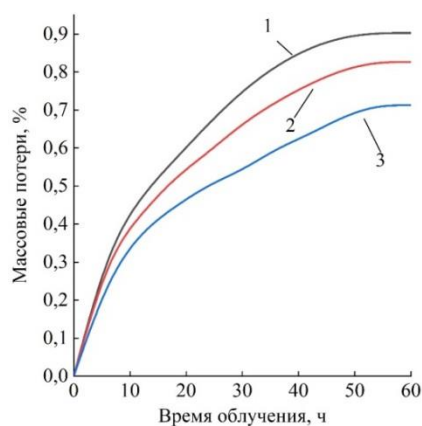


Рис.3. Массовые потери: 1 – ПИИ; 2 – ПЭТ; 3 – ФП

В таблице 1 представлены результаты измерений электрической ёмкости плёнок до и после облучения. Также в таблице приведены расчётные данные диэлектрической проницаемости. Относительную диэлектрическую проницаемость рассчитывали по следующей формуле:

$$\varepsilon = \frac{C \cdot d}{\varepsilon_0 \cdot S},$$

где ε - относительная диэлектрическая проницаемость; S - площадь контакта, составляющая 0,0004 м²; d - расстояние между контактами, составляющее 0,00002 м; ε_0 - электрическая постоянная, равная $8,85 \cdot 10^{-12}$ Ф/м; C - электрическая емкость, Ф.

Таблица 1. Электрические параметры полимеров

| Величины | Исследуемые образцы | | | | | |
|---|---------------------|------|------|------|------|------|
| | ПИ | | ПЭТ | | ФП | |
| | исх. | обл. | исх. | обл. | исх. | обл. |
| Электрическая ёмкость, нФ | 0,20 | 0,23 | 0,17 | 0,20 | 0,15 | 0,16 |
| Относительная диэлектрическая проницаемость | 1,15 | 1,29 | 0,96 | 1,15 | 0,84 | 0,90 |

Согласно таблице 1, воздействие ВУФ-облучения привело к увеличению электрических параметров всех образцов. Электрическая ёмкость увеличилась на 15, 17,6 и 6,7 % для полиимида, полиэтилентерефталата и фторопласта соответственно. Относительная диэлектрическая проницаемость увеличилась на 12,2, 19,8 и 7,1 % для полиимида, полиэтилентерефталата и фторопласта соответственно.

Увеличение значений электрических характеристик исследуемых полимеров предположительно связано с образованием свободных радикалов и разрушенных связей на поверхности плёнок после ВУФ-облучения.

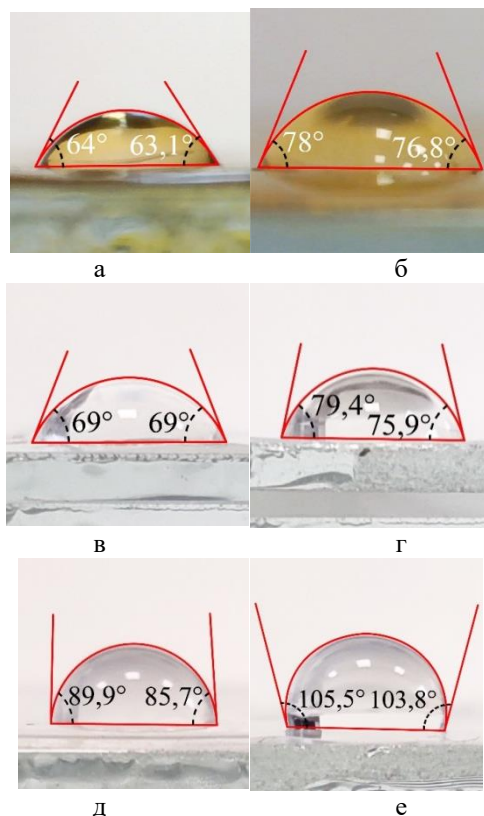


Рис.4. Изображение капли воды на исследуемых плёнках: а, б – ПИИ; в, г – ПЭТ; д, е – ФП; а, в, д – до облучения; б, г, е – после облучения

На рис. 4 представлены результаты анализа краевого угла смачивания исследуемых плёнок. До облучения все образцы имеют гидрофильные свойства. Краевой угол смачивания исходных плёнок ПЭТ и ФП выше, чем у полиимида на 8,6 и 38,2 % соответственно. Согласно рисунку, воздействие ВУФ-облучения на полиимидную плёнку привело к увеличению краевого угла смачивания на 21,8 %. Увеличение краевого угла смачивания говорит о гидрофобизации поверхности плёнки. На ПЭТ плёнку ВУФ-облучение повлияло также (краевой угол смачивания увеличился на 12,5 %. Поверхность плёнки также гидрофобизировалась. Однако, смачиваемые способности плёнок ПИ и ПЭТ остались гидрофильными. Что касается фторопластовой плёнки, ВУФ-облучение привело к увеличению краевого угла смачивания на 19,2 %. Смачиваемые способности ФП плёнки изменились с гидрофильных на гидрофобные.

Заключение

В результате проделанной работы можно сделать следующий вывод: воздействие вакуумного ультрафиолета на полимерные плёнки (полиимид, полиэтилентерефталат, фторопласт) в течение 60 часов сопровождается потерей массы полимеров. Массовые потери не превышают 0,9 %, наименьшие массовые потери при этом претерпела фторопластовая плёнка.

Вследствие воздействия ВУФ-облучения значения электрических параметров полимеров увеличиваются. Наибольшее увеличение электрической ёмкости и диэлектрической проницаемости наблюдается у ПЭТ плёнки. Предполагается, что такие изменения связаны с разрушением связей полимеров и образованием свободных радикалов на верхнем слое плёнок.

Вакуумный ультрафиолет изменяет поверхностные свойства полимерных плёнок. Так, смачиваемые способности полиимидной и полиэтилентерефталатовой плёнок остались гидрофильными, однако, краевой угол смачивания несколько увеличился. Поверхность фторопластовой плёнки до облучения характеризовалась гидрофильными свойствами, однако, после ВУФ-облучения поверхность плёнки приобрела гидрофобные свойства.

Исследование выполнено за счет гранта Российского научного фонда № 19-79-10064 (продление), <https://rscf.ru/project/19-79-10064/> с использованием оборудования на базе Центра высоких технологий БГТУ им. В.Г. Шухова.

Список литературы

1. Lu W. Impact of Vacuum Ultraviolet (VUV) Intensity on the Performance of VUV Irradiation for the Mineralization of Trimethoprim in Neutral Media: Kinetics, Mechanisms, and by-Products Formation / Lu W., Ren S., Zhang Y., Wen X., Zhang Z., Wang A. // Environmental Science and Pollution Research. - 2023. - V. 30. - № 57. - P. 1-15.

2. Михайлов М. М. Оптические свойства и радиационная стойкость полипропилена, модифицированного наночастицами MgO / М. М. Михайлов, В. А.

Горончко // Космические аппараты и технологии. - 2022. - Т. 6. - № 2(40). - С. 102-108.

3. Мьяленко Д. М. Влияние термического, радиационно-химического и фотометрического воздействия на деструкцию и "старение" полимерных материалов / Д. М. Мьяленко // Актуальные вопросы молочной промышленности, межотраслевые технологии и системы управления качеством. - 2020. - Т. 1. - № 1(1). - С. 406-411.

4. Miyaji Y. A Study on the Estimation of RIC (Radiation-Induced Conductivity) of Proton Irradiated Polyimide / Miyaji Y., Miyake H., Tanaka Y. // Polymers. - 2023. - V. 15. - P. 337.

5. Гасанов С. К. Использование вакуумного ультрафиолета для уменьшения шероховатости поверхности полимерных композитов / С. К. Гасанов, Р. Н. Ястребинский, В. И. Павленко // Успехи современного естествознания. - 2015. - № 10. - С. 11-14.

6. Федотова О. Б. Исследование изменения структуры полиэтиленового слоя упаковки, контактирующего с пищевым продуктом, при воздействии ультрафиолетового излучения / О. Б. Федотова, Н. С. Пряничникова // Пищевые системы. - 2021. - Т. 4. № 1. - С. 56-61.

7. Черкашина Н. И. Исследование влияния вакуумного ультрафиолета на морфологию поверхности нанонаполненных полимерных композиционных материалов в условиях, приближенных к условиям околоземного космического пространства / Н. И. Черкашина, В. И. Павленко, А. С. Едаменко, П. В. Матюхин // Современные проблемы науки и образования. - 2012. - № 6. - С. 129.

8. Черкашина Н. И. Воздействие вакуумного ультрафиолета на полимерные композиты терморегулирующего назначения / Н. И. Черкашина // Международный научно-исследовательский журнал. - 2016. - № 7-4(49). - С. 72-77.

СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ ПОВРЕЖДАЕМОСТИ ПОВЕРХНОСТНОГО СЛОЯ ПЛАСТИН ТАНТАЛА ПРИ ИМПУЛЬСНОМ ВОЗДЕЙСТВИИ ПОТОКОВ БЫСТРЫХ ИОНОВ И ВЫСОКОТЕМПЕРАТУРНОЙ ПЛОТНОЙ ПЛАЗМЫ ГЕЛИЯ И ДЕЙТЕРИЯ В УСТАНОВКАХ «ПЛАЗМЕННЫЙ ФОКУС» С РАЗЛИЧНОЙ ЭНЕРГЕТИКОЙ

*Дёмин А.С., Епифанов Н.А., Морозов Е.В.
Федеральное государственное
бюджетное учреждение науки
Институт металлургии и материаловедения
им. А.А. Байкова Российской академии наук*

Аннотация

В ходе работы были изучены последствия импульсного пучково-плазменного воздействия потоков ионов гелия, дейтерия и их высокотемпературной плазмы, возникающие в процессе Z-пинчирования токо-плазменного шнура на установках типа Плаз-

менный фокус (с энергией, запасенной в конденсаторах: ПФ-1000 Е ≈ 1МДж и ПФ Вихрь Е ≈ 5 кДж) на образцы из тантала. Выявлены схожие черты в изменениях текстуры, микроструктуры, микротвердости поверхностных слоев образцов тантала после воздействия плазмы и ионов, а также их различия. Произведено математическое моделирование физических процессов, происходящих в материале и на его поверхности в условиях, соответствующих эксперименту.

Изучены особенности и общие черты изменений рельефа, микроструктуры, микротвердости поверхностного слоя под воздействием двух разных газов гелия и дейтерия, которые напускались в установку ПФ Вихрь в качестве рабочего газа. Явных различий в рельефе и морфологии поверхностного слоя не выявлено, но микротвердость значительно возросла после воздействия гелиевой плазмы и ионов по сравнению с дейтериевой плазмой и ионами.

Введение

Тантал имеет технологический потенциал в промышленности, где требуется работа при высоких температурах и механических нагрузках, например, в ядерной и термоядерной энергетике [1]. Тантал обладает быстрым спадом наведенной активности, что косвенно позволяет считать его малоактивируемым материалом [2] и применять в узлах регулирования ядерных реакторов на быстрых нейтронах [3, 4].

В исследованиях [5, 6] тантал предлагается использовать как промежуточный слой между вольфрамом (который в проекте DEMO рассматривается как материал первой стенки) и малоактивируемой ферритно-мартенситной сталью, применяемой как конструкционный материал. Данный слой тантала снижает напряжения, возникающие между вольфрамом и сталью [7]. Изучается возможность использования вольфрама, тантала как материалов нейтронпроизводящих мишеней в источниках нейтронов [8].

Поведение тантала под воздействием высоких температур, радиационных воздействий, значительных механических микронапряжений и влияние газообразных элементов-трансмутантов на тантал требует изучения.

Целью данной работы было изучение влияния пучково-плазменного воздействия, возникающего в установках Плазменный фокус, на механические свойства, изменения структуры в поверхностном слое тантала.

Методика эксперимента

Микротвердость определяли на микротвердомере Shimadzu DUN-211S (Япония) методом восстановленного отпечатка алмазным индентором Виккерса. При нагрузках $F_{\max} = 50$ гс и $F_{\max} = 16$ гс, времени выдержки 5 с. Микротвердость рассчитывалась по формуле $HV = 189,1 \cdot F_{\max} / L^2$, где F_{\max} – максимальная нагрузка, мН; L – длина диагонали восстановленного отпечатка, мкм. Относительная погрешность результатов измерений микротвердости составляет около 6% при доверительной вероятности $P = 0,95$. Пере-

вод значений микротвердости из единиц по шкале Виккерса (HV) в МПа выполнялся умножением на коэффициент 9,807.

Регистрация рентгеновских дифракционных спектров исходных образцов и образцов после радиационного воздействия проведена в $\text{CuK}\alpha$ – излучении на приборе SHIMADZU XRD-6000. Интервал углов съёмки $2\theta = (36,0 - 123,0)$ град., шаг 0,04 град., скорость 3 град/мин.

Облучение образцов происходило на двух различных установках типа Плазменный фокус: ПФ-1000 и ПФ Вихрь.

Эксперимент на установке ПФ-1000 проводился в атмосфере дейтерия с давлением ~300 Па. В анодной зоне рабочей камеры установки ПФ-1000 с помощью алюминиевого держателя зажимались края пластины тантала размером $100 \times 90 \times 0,20$ мм. Расстояние от пластины до анода по оси Z составляло 15 см (рис. 1а). Для дальнейшего исследования из облученной пластины изготавливались образцы размером 10×10 мм. Поскольку зона облучения имеет значительно больший диаметр, чем вырезаемый образец и имеет спадающий градиент интенсивности воздействия от центра к краю, образцы брались из характерных зон с максимальной интенсивностью воздействия и с «периферии», где плотность мощности была минимальна. Максимальная плотность мощности быстрых ионов дейтерия составляла $q \sim 10^{11}$ Вт/см² при длительности импульса $\tau \sim 100$ нс. Количество импульсов воздействий N на каждую из двух Та пластин составляло $N = 4$.

Образцы для облучения в установке ПФ Вихрь имели размер $15 \times 15 \times 0,20$ мм, располагались так же в катодной зоне материаловедческой камеры в 4 см от анода по оси Z (рис. 1б). В эксперименте изменялось количество импульсов от образца к образцу от 12 до 30. Так же изменялся рабочий газ: для одних образцов использовался дейтерий, для других гелий. Максимальная плотность мощности быстрых ионов дейтерия и гелия составляла $q \sim 2 \cdot 10^9$ Вт/см², $\tau \sim 20$ нс.

Плотность мощности импульсных потоков плазмы в каждой из установок отличается в меньшую сторону, примерно на порядок, но в то же время, обладает большей длительностью воздействия, по сравнению с потоками быстрых ионов.

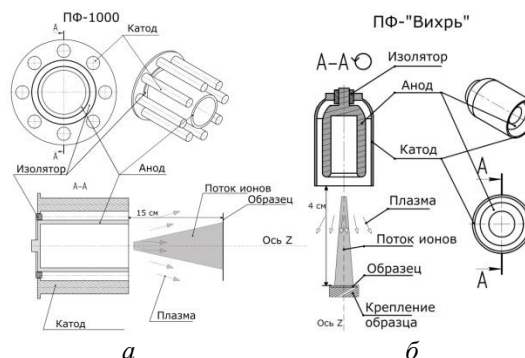


Рис.1. Схема облучения тантала: а – в установке ПФ-1000, б – в установке ПФ Вихрь

Результаты экспериментов

Произведено математическое моделирование взаимодействия быстрых ионов дейтерия и гелия на образец из тантала в условиях, соответствующих эксперименту, по методике, предложенной в работе [9].

Результаты вычислений показали, что на более мощной установке ПФ-1000 происходит испарение поверхностного слоя ~6 мкм, это превышает глубину проникновения ионов дейтерия и гелия с энергией 100 кэВ для танталовой мишени (~1 мкм, ~0.5 мкм соответственно). Однако на установке ПФ-Вихрь толщина испаренного слоя ~0,5 мкм, что меньше глубины проникновения данных ионов. Эта разница является причиной более выраженных изменений микроструктуры в результате облучения на установке ПФ Вихрь (появляется множество микротрещин, кратеров), что является следствием образования дополнительных напряжений, создаваемых этими газами и термических напряжений, образующихся на стадии снижения температуры.

После воздействия потоками ионов дейтерия и дейтериевой плазмы на установке ПФ Вихрь, поверхность образцов имеет развитый волнообразный рельеф, имеются кратеры (результат выхода газов), микротрещины (рис. 2б), в отличие от изменений поверхностного слоя после облучения на установке ПФ-1000 (рис. 2а), где волнообразный рельеф менее развит, а кратеры и микротрещины отсутствуют.

Это может быть связано с разницей в толщине испаренного и оплавленного слоя.

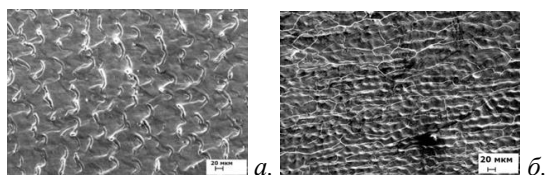


Рис. 2. Типичный характер морфологии образцов облученных на установках, а – ПФ-1000 и б – ПФ «Вихрь»

Ионы газов имплантируются в поверхностный слой материала образца, что приводит к увеличению микротвердости поверхности образцов, облученных ионами гелия или дейтерия в экспериментах на ПФ Вихрь. На установке ПФ-1000 после облучения дейтерием микротвердость или не существенно меняется в меньшую сторону, или не изменяется вовсе (рис. 3).

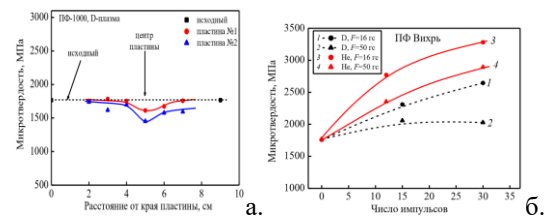


Рис. 3. Изменение микротвердости поверхностного слоя. а – двух одинаковых пластин Та после их облучения дейтериевой плазмой в ПФ-1000 при одинаковых условиях эксперимента, б - облученных в ПФ Вихрь с различным числом импульсных воздействий дейтериевой (кривые 1, 2) и гелиевой (кривые 3, 4) плазмы (F – нагрузка на инденторе)

Образец Та имеет холоднодеформированную текстуру, сформировавшуюся при прокатке, что выражается в повышении интенсивности рефлексов от семейства плоскостей $\{200\}$ и в меньшей степени $\{211\}$ в сравнении с интенсивностью пиков для стандартного образца Та в отожженном состоянии (см. рис. 4а и 4и). После облучения текстура во всех случаях меняется, наиболее заметно – в образцах, облученных в ПФ Вихрь (рис. 4д-з). В них наблюдается увеличение интенсивности рефлексов от семейства плоскостей $\{211\}$, и в то же время, рефлексы от плоскостей $\{110\}$, наиболее интенсивные для стандартного образца (рис. 4и), напротив, заметно ослабевают. Различий в текстуре в зависимости от типа используемого рабочего газа (дейтерий или гелий) не обнаружено. Образцы, вырезанные из разных участков широкой пластины, облученной на ПФ-1000, имеют примерно одинаковую текстуру с более интенсивными по сравнению с эталоном рефлексами от плоскости $\{211\}$, а также от плоскостей $\{200\}$ и $\{222\}$ (рис. 4б – з). Формирование текстуры, вероятнее всего, связано с направленным теплоотводом при кристаллизации поверхностного слоя, оплавленного при облучении в ПФ.

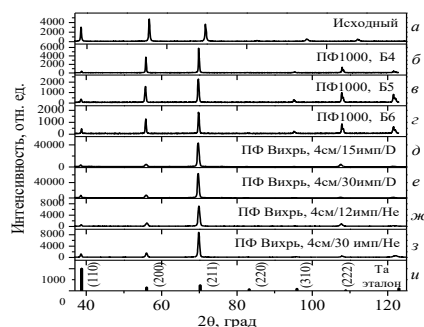


Рис. 4. Рентгенограммы образцов Та в исходном состоянии (а) и облученных в установках ПФ-1000 (б – з) и ПФ-Вихрь (д – з). би-в отожженном состоянии. Б4 - Б6 вырезки из центральной зоны пластины, облученной на установке ПФ-1000 (рис. б – з)

Обсуждение результатов

Проведено облучение образцов тантала на двух различных по характеристикам установках типа Плазменный фокус с различными параметрами импульсного пучково-плазменного воздействия.

Исходя из результатов вычислений, толщина испаренного слоя за 1 импульс на образцах тантала в реализованном режиме облучения для установки ПФ-1000 превышала глубину пробега ионов дейтерия с энергией 100 кэВ, а для установки ПФ Вихрь была меньше. В то же время глубина проплавления за одно воздействие на установке ПФ-1000 значительно превышала глубину проплавления поверхностного слоя на установке ПФ Вихрь.

При облучении образцов на установке ПФ 1000 формируется структура с размером ячеек ≤ 200 нм в результате сверхбыстрого плавления и кристаллиза-

ции в приповерхностных слоях. На образцах, облученных на установке ПФ Вихрь, присутствует множество кратеров (результат выхода имплантированных газов). Эти отличия связаны с разницей в величине испаренного слоя и скорости охлаждения поверхностного слоя после воздействия. На установке ПФ 1000 происходило полное испарение слоя с имплантированными ионами, в то время, как на установке ПФ Вихрь имплантированный дейтерий или гелий перераспределялись в жидкой фазе, вызывая в ней процессы «кипения» и создавая дополнительные напряжения.

Исследовано влияние импульсного ионно-плазменного воздействия на микротвердость поверхности образцов тантала. Обнаружено, что при условиях облучения, которые создаются на установке ПФ 1000, изменения микротвердости незначительны. В то же время при более мягком режиме на установке ПФ Вихрь микротвердость поверхностного слоя возрастает.

Произведено сравнение изменений рельефа, микроструктуры, микротвердости поверхностного слоя под воздействием двух разных газов гелия и дейтерия, которые напускались в установку ПФ Вихрь в качестве рабочего газа. Явных различий в рельефе и морфологии поверхностного слоя не выявлено, но микротвердость значительно возросла после воздействия гелиевой плазмы и ионов по сравнению с дейтериевой плазмой и ионами. При этом с увеличением числа импульсных воздействий в исследованном диапазоне (до $N=30$) эффект упрочнения переплавленного слоя растет, причем более существенно при использовании гелия.

Список литературы

1. Конструкционные материалы ядерных реакторов. В 2-х частях. Ч. II. Структура, свойства, назначение. Под общей редакцией Н.М. Бескорвайного. М.: Атомиздат. – 1977. – 256 с.
2. Калинин Б.А., Польский В.И., Якушин В.Л., Чернов И.И. Материаловедческие проблемы экологии в области ядерной энергетики // М.: НИЯУ МИФИ. – 2010. – 184 с.
3. Murgatroyd R.A., Kelly B.T. Technology and assessment of neutron absorbing materials // Atomic Energy Rev. – 15. – 1. – 1977.
4. Е.П.Клочков, Е.М.Муралева Использование тантала для органов регулирования ядерных реакторов на быстрых нейтронах. Известия Самарского научного центра Российской академии наук // 2012. – т.14. – №4(4). – С. 975-979.
5. D. Bachurina, A. Suchkov, J. Gurova, V. Kliucharev, V. Vorkel, M. Savelyev, P. So-mov, O. Sevryukov. Brazing tungsten/tantalum/RAFM steel joint for DEMO by fully reduced activation brazing alloy 48Ti-48Zr-4Be // Metals (Basel) . – 2021. – 11. – P. 1417. doi: 10.3390/met11091417
6. Y. Gasparyan, D. Bachurina, V. Efimov, J. Gurova, F. Podolyako, N. Sergeev, I. Sorokin, A. Suchkov, N. Bobyr, I. Kozlov, E. Kulikova, A. Spitsyn Deuterium retention in the elements of plasma facing components for the DEMO first wall // Journal of Nuclear Materials. – 567. – 2022. – 153837
7. D. Bachurina, A. Suchkov, B. Kalin, O. Sevriukov, I. Fedotov, P. Dzhumaev, A. Ivannikov, M. Leont'eva-Smirnova, E. Mozhanov. Joining of Tungsten with Low-Activation Ferritic-Martensitic Steel and Vanadium Alloys for Demo Reactor // Nucl. Mater. Energy. – 2018. – 15. – P. 135–142.
8. Thak Sang Byun, Stuart A. Maloy Dose dependence of mechanical properties in tantalum and tantalum alloys after low temperature irradiation. Journal of Nuclear Materials. – 377. 2008. pp. 72–79.
9. Грибков В.А., Латышев С.В., Масляев С.А., Пименов В.Н. Численное моделирование взаимодействия импульсных потоков энергии с материалом в установках Плазменный фокус Физика и химия обработки материалов. – 2011. – №6. – С. 16-22.

ВОЗДЕЙСТВИЕ ИМПУЛЬСНЫХ ПОТОКОВ ВЫСОКОТЕМПЕРАТУРНОЙ ПЛАЗМЫ И БЫСТРЫХ ИОНОВ НА НИКЕЛЕВЫЙ СПЛАВ ИНКОНЕЛЬ 718

Морозов Е.В., Дёмин А.С., Епифанов Н.А.

Федеральное государственное бюджетное учреждение науки

Институт металлургии и материаловедения им. А.А. Байкова Российской академии наук

В подавляющем большинстве случаев, воздействие пучков частиц высоких энергий, влияет на внешние слои исследуемых материалов. Наиболее ярко, подобное влияние прослеживается в условиях, когда жёсткие радиационно-термические импульсные потоки частиц, прогревают поверхностный слой выше порога плавления. В современной авиакосмической промышленности, всё более значительную роль начинают играть аддитивные технологии, в частности, в сфере жаропрочных металлов, таких, как Инконель 718, обладающим сложным, многокомпонентным химическим составом: Ni=52,50%, Cr=19,00%, Fe=18,86, Mo=3,00%, Al=0,50%, Ti=0,90%, Nb=5,10%, C< 0,08%, B< 0.06%.

Это сплав обладает нужной комбинацией физико-химических свойств, которые сохраняет даже при достижении температуры 800°C и применяется в области реактивной техники, турбинах и т.д.

При производстве подобных материалов, важнейшим этапом является оценка температурно-временных характеристик сплава, а так же оценка структурно-фазового состояния получаемого материала.

Ранее, было проведено[4] комплексное изучение многократного импульсного воздействия пучков высокоэнергетических ионов и плазмы на материал Инконель 718, изготовленного по аддитивной технологии, в мягком режиме.

Более жёсткие воздействия ионов и плазмы на Инконель 718, приводящие к распылению и пере-

правлению поверхностных слоёв, были изучены ранее в статье [1].

В текущем докладе рассматривается мощное импульсное воздействие ионов гелия ($E \sim 100 \text{ кДж}$) и гелиевой плазмы ($E \sim 1 \text{ кДж}$), генерируемой в установке Плазменный фокус на структуру и фазовое состояние, приготовленного по аддитивной технологии, методом селективного лазерного спекания, сплава Инконель 718, подвергнутого так же, термической обработке – закалке и старению. Так же, проводились эксперименты по модификации структуры материала, пучково-плазменным воздействием. В работе применялись методы оптической микроскопии (Neophot 32). Рентгенофазовый анализ поверхностного слоя проводился с помощью рентген-дифрактометра Ultima IV Rigaku (Япония). Микротвердость $HV_{0,05}$ измерялась прибором Shimadzu DUN-211S (Япония), с алмазным индентором Виккерса.

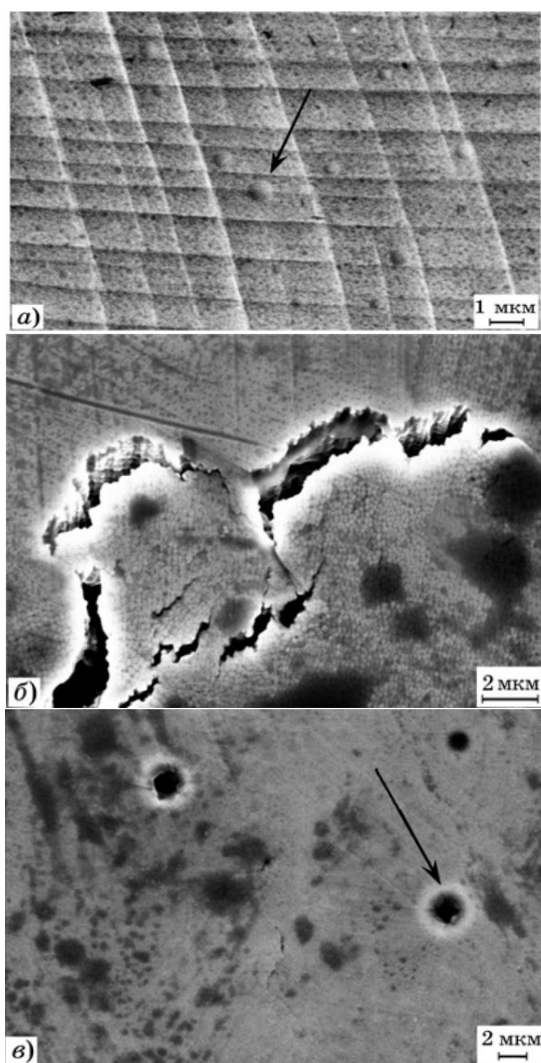
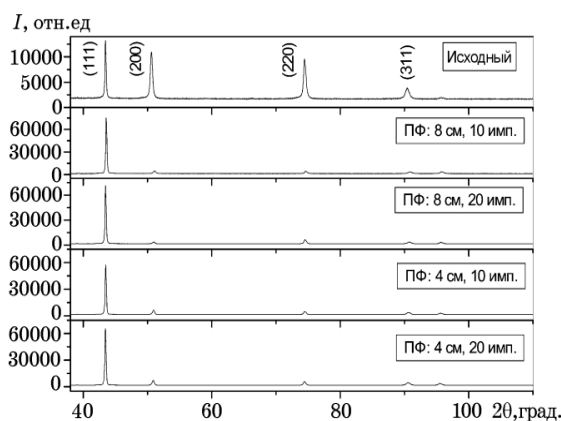


Рис.1. Микроструктура (СЭМ) участков центральной зоны поверхности сплава Инконель 718, облученного потоками ИГ и ГП в мягком режиме с числом импульсных воздействий $N = 5$ (а) и в жёстком режиме с $N = 10$ (б) и 20 (в). Стрелками отмечены блистеры с неразрушенными (а) и разрушенными (в) оболочками

Влияние разных режимов облучения на структуру материала.

Статья [1], отмечает наличие пластических деформаций в мягких режимах облучения, о чём свидетельствуют многочисленные линии скольжения на отдельных участках поверхности образцов. В более жёстком режиме облучения, линии скольжения выражены слабее (рис.1, а—в). Это может указывать на уменьшение данного процесса, что связано с иным, более интенсивным характером повреждения поверхностного слоя, таким как более разнообразный рельеф, микротрещины, осаждённые элементы анода. установки (темные пятна на рис.1, б, в), а также, ярко выраженный процесс блистеринга с последующего разрушения оболочек (рис.1, в).



Параметр решетки a , Å, твердого раствора в образцах сплава Инконель 718 до и после облучения

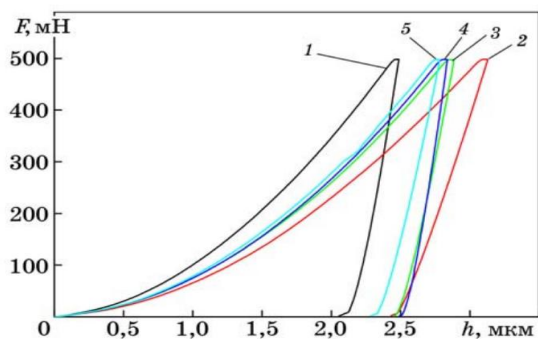
| Исходный | Условия облучения в установке ПФ | | | |
|----------|----------------------------------|------------------|------------------|------------------|
| | 8 см, 10 имп. | 8 см, 20 имп. | 4 см, 10 имп. | 4 см, 20 имп. |
| 3,600(7) | 3,596(7) | 3,600(3) | 3,602(3) | 3,605(6) |

Рис.2. Рентгеновские дифрактограммы и параметры решётки образцов сплава Инконель 718

Проводилось рентгеноструктурное исследование образцов материала Инконель 718, до и после облучения на установке Плазменный фокус (рис.2). Из дифрактограмм видно, что структура образца является однофазным твёрдым раствором с ГЦК решёткой, на основе системы Ni-Cr-Fe с преобладанием никеля (>50%) в котором растворены прочие компоненты. Одной из особенностей, выявленной до облучения, стала текстура, выраженная в ориентации кристаллитов по плоскостям (220) проявившаяся в увеличении интенсивности рефлекса (220). После облучения, ориентация кристаллитов переменялась в сторону упорядочивания по плоскостям (111), предполагается, что это произошло вследствие изменения теплоотвода при кристаллизации оплавленной поверхности. Обычно, в ГЦК - сплавах, скольжение идет по плоскостям (111) в направлении с более плотной упаковкой [2]. Как правило, пластическому течению и наблюдаемой блочной структуре (см. рис.1, а).

способствует возникновение текстуры в поверхностном слое материала, что наблюдалось так же при воздействии потоков ионов и плазмы на медный сплав в медном сплаве Cu-10Ga-4Ni в работе[3].

Так же, в работе представлены результаты измерения параметра a , которые можно видеть в таблице. Видно, что в мягком (с расстоянием от образца до анода $L = 8$ см) режиме, параметр a ниже исходного, когда как жёстком (с расстоянием от образца до анода $L = 4$ см), он наоборот выше. Можно видеть, что с ростом количества импульсов, параметр увеличивается, что Это определяется влиянием ряда факторов. Например, уменьшение межплоскостных расстояний, вызванных импульсным воздействием[4]. Так же на этот параметр может повлиять испарение из поверхностного слоя примесных элементов, расположенных между узлами кристаллической решетки (O_2 , C и др.). Одновременно с этим, происходит имплантация атомов рабочего газа в кристаллическую решётку поверхностного слоя образца, что должно увеличивать период решётки. Результирующий параметр решётки a , является соотношением двух этих факторов. По мере приближения образца к аноду установки, возрастает плотность мощности, что приводит к более интенсивной имплантации рабочего газа в поверхностный слой, таким образом, фактор роста параметра a становится основным. Предел погрешности измерений параметра решётки a колебался в пределах ± 3 , ± 6 и ± 7 в зависимости от образца.



Микротвердость ПС образца сплава Инконель 718 в исходном состоянии и после воздействия потоков ионов гелия и гелиевой плазмы в мягком ($L = 8$ см) и жестком ($L = 4$ см) режимах облучения

| Показатель | Исходный | Условия облучения в установке ПФ | | | | Без оплавления поверхности [20] |
|------------------------------|----------|----------------------------------|---------------|---------------|---------------|---------------------------------|
| | | 8 см, 10 имп. | 8 см, 20 имп. | 4 см, 10 имп. | 4 см, 20 имп. | |
| $HV_{0,05}$, МПа | 3920 | 2620 | 2660 | 2720 | 2760 | 3310 |
| h_{max} , мкм ² | 2,47 | 3,18 | 2,99 | 2,83 | 2,90 | — |

²Максимальная глубина вдавливания индентора.

Рис. 3. Типичные кривые индентирования образцов сплава Инконель 718 до (кр. 1) и после воздействия потоков ИГ и ГП в мягком ($q = 2 \cdot 10^8$ Вт/см², $\tau = 50$ нс, кр. 2 — $L = 8$ см, 10 имп.; кр. 3 — 8 см, 20 имп.) и жестком ($q = 1,5 \cdot 10^9$ Вт/см², $\tau = 25$ нс, кр. 4 — $L = 4$ см, 10 имп.; кр. 5 — 4 см, 20 имп.) режимах облучения в установке ПФ. Так же представлены результаты численного моделирования давления ударной волны

Результаты измерений приведены в таблице, как для исходного, так и для облучённого состояния. Из таблицы видно, что многократное импульсное воздействие на образец, ведёт к уменьшению изначальной величины $HV_{0,05}$ на 30%, как жёстком, так и в мягком режиме облучения, что значительно выше, чем снижение величины микротвёрдости, при использовании в качестве рабочего газа дейтерия[5]. Вероятнее всего, это связано с температурным воздействием импульсных потоков, приводящие к резким скачкам температуры выше порога плавления и последующей высокоскоростной кристаллизацией поверхностного слоя материала, что и приводит к снижению значения $HV_{0,05}$, достигнутого методом селективного лазерного спекания. Так же, приведены кривые индентирования образцов (рис.3) в обоих режимах. Максимальная глубина вдавливания индентора в образец, по расчетам $L_{пл} = 1,6$ мкм [1], в то время как на практике, примерно в два раза больше, $h_{max} \approx 3$ мкм. Глубина индентирования затрагивала как переплавленный поверхностный слой, так и находящийся под ним в зоне термического влияния, но не плавившийся более глубокий, слой. Характер графика «нагрузка—разгрузка» (см. рис. 3) говорит о том, что упругие свойства отмеченных слоев близки для обоих режимов облучения.

Выводы

1. Проведено модельное облучение никелевого сплава Инконель 718 приготовленного по аддитивной технологии селективного лазерного спекания в установке Плазменный фокус, в двух режимах облучения: мягкого (плотность мощности излучения $q = 2 \cdot 10^8$ Вт/см² при длительности импульса $t = 50$ нс) и жесткого ($q = 1,5 \cdot 10^9$ Вт/см², $t = 25$ нс).
2. Микроструктура поверхности показала наличие линий скольжения, микротрещин, наплывов и блистеринга на поверхности образцов после облучения.
3. Рентгеноструктурное исследование образцов сплава показало изменение ориентации кристаллитов и интенсивности рефлекса в поверхностном слое, после облучения.
4. Параметр кристаллической решётки a изменялся под действием облучения, что связано, преимущественно с имплантацией рабочего газа установки в поверхностный слой.
5. Уменьшение микротвёрдости сплава связано с температурным воздействием импульсных потоков, приводящие к резким скачкам температуры выше порога плавления и последующей высокоскоростной кристаллизацией поверхностного слоя материала.

Список литературы

1. Боровицкая, И.В. Повреждаемость поверхностного слоя сплава Инконель 718 импульсными пучково-плазменными потоками / И.В. Боровицкая, А.С. Демин, С.В. Латышев, С.А. Масляев, И.С. Монахов, Е.В. Морозов, В.Н. Пименов, И.П. Сасиновская, Г.Г. Бондаренко, А.И. Гайдар // ФХОМ. 2023. №2. С. 5—17.

2. Бондаренко, Г.Г. Радиационная физика, структура и прочность твердых тел / Г.Г. Бондаренко. — М. : Изд-во Лаборатория знаний, 2016. 462 с.

3. Боровицкая, И.В. Влияние облучения высоко-температурной импульсной дейтериевой плазмой на структуру и механические свойства поверхности сплавов систем Cu-Ga и Cu-Ga- Ni / И.В. Боровицкая, В.Н. Пименов, С.А. Масляев, А.Б. Михайлова, Г.Г. Бондаренко, Е.В. Матвеев, А.И. Гайдар, М. Падух, А.С. Дёмин, Н.А. Епифанов, Е.В. Морозов // *Металлы*. 2022. №1. С.55—64.

4. Перлович, Ю.А. Изменение структуры и текстуры в объеме оболочечных труб из сплавов на основе циркония при ионно-плазменной обработке поверхности / Ю.А. Перлович, М.М. Грехов, М.Г. Исаенкова, В.А. Фесенкоф, Б.А. Калинин, В.Л. Якушин // *Вопр. атомной науки и техники. Сер. Физика радиационных повреждений и радиационное материаловедение*. 2004. Т.85. №3. С.59—65

5. Боровицкая И.В., Грибков В.А., Григорович К.В., Демин А.С., Масляев С.А., Морозов Е.В., Пименов В.Н., Спрыгин Г.С., Цепелев А.Б., Гусаков М.С., Логачев И.А., Бондаренко Г.Г. Гайдар., А.И. Воздействие импульсных потоков ионов гелия и гелиевой плазмы на сплав Инконель 718. // *Металлы*, 2018. № 5. С.39—47.

РАЗРАБОТКА ПАРАМЕТРИЗИРОВАННОГО 3D ПРИНТЕРА

Галиахмедов Э.И., Гончаров Н.П., Герасимов Р.А., Нгуен А., Тулицкий Д.С.
**Национальный исследовательский университет «Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ**

Аннотация

Данная работа заключается в полной доработке и усовершенствовании предыдущей версии ПО и моделированием новых деталей к 3D принтеру. ПО способно изменять размеры и характеристики деталей принтера под требуемые размеры. При параметризации будет воспроизводиться параметризация деталей, которые были смоделированы в программе "Компас-3D".

Введение

В реалиях нашего мира потребность в воссоздании требуемых деталей и компонентов в машиностроении, медицине, учёбе и ремонте становится всё более и более важной частью. На помощь приходит 3D печать, что значительно уменьшает затраты на материал и время. Частой проблемой в этих отраслях является проблема параметризации требуемых компонентов. Наша работа основывается на настройке размеров и различных характеристик 3D принтера.

Приложение работает по принципу изменения размеров компонентов 3D Принтера под требуемые,

что значительно улучшает качество сборки, уменьшает потраченное время и увеличивает круг потенциальных заказчиков.

Аналоги разрабатываемой системы

В данное время наша программа не имеет каких-либо хороших и способных аналогов, которые могли бы параметризовать принтер под требуемые задачи. Мы в свою очередь предоставляем регулировку размеров и рабочей области 3D принтера. Есть лишь попытки создать аналоги в виде скриптов и макросов. Макрос – это набор команд, записанных для автоматического выполнения, как правило, не очень сложных действий в приложениях. Программа – это независимое приложение, которое может выполнять широкий спектр функций, включая расчёт сложных параметров при построении сборки. Именно поэтому наша программа существенно опережает все попытки в виде макросов и скриптов. Существующие модели 3D принтеров многообразны и применяются в различных сферах, однако возникают затруднения с выбором идеальной опции из-за жестких стандартов. Например, приобретение среднего по размеру принтера может быть неоправданным, если требуется печатать только мелких деталей. Однако, представьте себе возможность предварительной настройки и указания параметров для получения оптимально подходящего принтера с необходимыми размерами. Это позволяет сэкономить ресурсы и материалы, получая идеально подходящий 3D принтер для конкретных задач.

Среда разработки

В качестве среды разработки мы решили использовать нашу отечественную систему под названием Компас-3D [1-2], которая является одной из самых продвинутых и инновационных. Программа позволяет создать изделие любой сложности и любого размера. А также есть возможность проверить сможет ли конструкция быть собрана, и как это будет выглядеть в сборном виде. Немаловажной особенностью является создание макетов и чертежей, которые можно оформить по заданным ГОСТам.

Главной особенностью системы КОМПАС-3D является обеспечение сквозного проектирования начиная с первоначального проектирования идеи в 3D до полной реализации проекта в действительности. В связке с САМ-системами появляется возможность уменьшить затраченное время для подготовки компонента. В изучении КОМПАС-3D не представляет трудностей, интерфейс достаточно простой и сам помогает новичкам.

Разработка 3D моделей

Процесс изменения характеристик принтера подразумевает настройку его величин в соответствии с требуемыми размерами рабочей зоны. С помощью этого мы получаем требуемый размер принтера, который подходит под требования заказчика. Это может быть возможным с помощью параметризации множества деталей, из которых состоит принтер. Но для того, чтобы настроить параметры деталей, преж-

де всего их нужно смоделировать и разработать в программном обеспечении Компас-3D.

Непосредственно сборка принтера включает в себя более ста неповторимых деталей, большинство из которых могут быть настроены в соответствии с входными данными. Среди них: держатели валов, ролики, провода, рельсы, зубчатые шестерни. Все эти элементы принтера представлены в форме трехмерных моделей, разработанных в соответствии с соответствующей документацией. Каждая из этих деталей, которые используются в процессе автоматической параметризации, изменяются при задании различных величин. Например, длина рельсов зависит от площади рабочей зоны, а количество необходимых держателей валов прямо связано с длиной самих валов.

Перечень деталей, которые параметризуются в программе:

- Профиль конструкционный 20x20SH;
- Замкнутый ремень;
- Ремень для каретки;
- Винт М3;
- ШВП SFU1610 винт (Вал);
- Алюминиевый лист толщиной 5мм;
- Боросиликатное стекло толщиной 4мм.

Программная часть

Созданное программное обеспечение позволяет перестраивать готовую сборку по заданным пользователем параметрам. Таким образом, реализуется возможность получать несколько готовых моделей принтера с различными параметрами за короткий промежуток времени. При разработке использовалась технология API Компас-3D от российской компании Аскон, являющаяся связующим звеном между программой для проектирования трехмерных моделей Компас-3D и написанной на языке программирования С# программным обеспечением. На рис.1 представлена архитектура программного обеспечения.

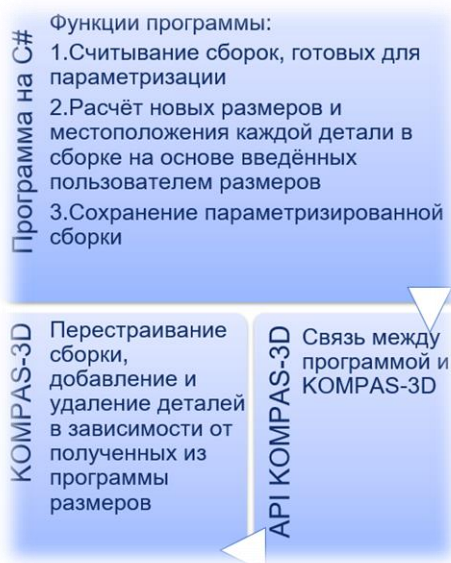


Рис.1. Архитектура программного обеспечения

При начале работы программы пользователю предоставляется возможность выбрать готовую параметризирующуюся сборку (на данный момент реализована только одна) и открыть её в Компас-3D. Далее появятся текстовые поля, в которые можно будет вписать целочисленные параметры для перестройки принтера (рис.2 и рис.3). Специальные ограничения не дадут пользователю перестроить принтер с недопустимыми габаритами (рис.4). Для ширины и длины рабочей области этот интервал составляет от 150 до 500 миллиметров из-за особенностей используемого двигателя. Только когда все данные подойдут под заданный интервал, кнопка "Параметризовать" станет активной.

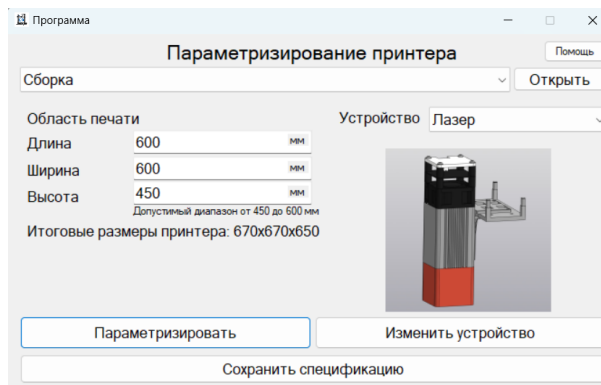


Рис.2. Интерфейс программы (1)

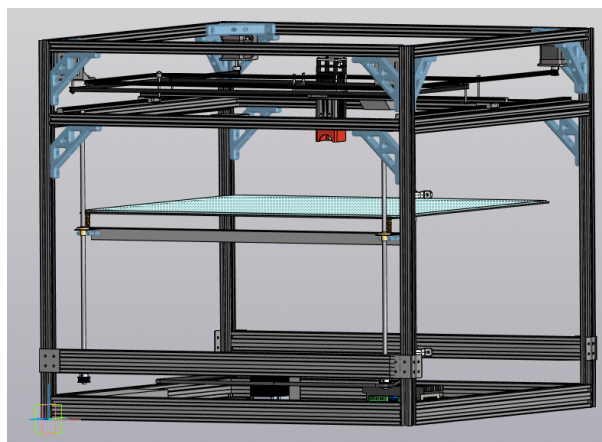


Рис.3. 3D модель принтера после параметризации

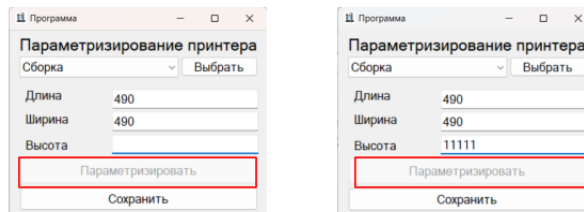


Рис.4. Недопустимые параметры

Программное обеспечение разделено на модули, каждый из которых отвечает за одну из частей параметризации: взаимодействие с программой Компас-

3D, параметризация отдельных компонентов принтера (в том числе сложные детали, требующие отдельной работы с ними). Так, например, параметризация ремня (рис.5) реализуется в отдельном от сборки окне, где с помощью 6 опорных точек изменяется его геометрия под новые размеры принтера. Получение и изменение координат и других параметров деталей происходит через их уникальный числовой идентификатор.

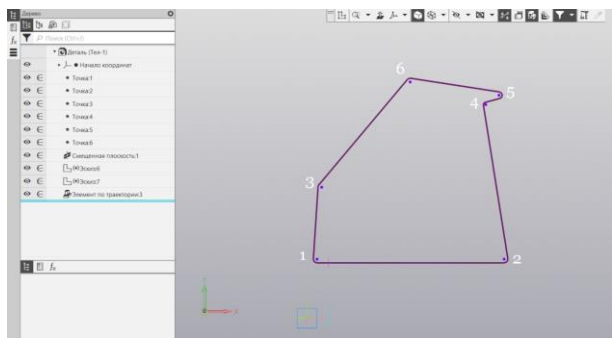


Рис.5. Ремень, перемещающий стол по оси Z

Разработка программного обеспечения ведётся в Visual Studio с разнообразным набором инструментов для создания приложений на разных языках программирования, таких как используемый в проекте C# [3]. Интерфейс программы разработан с использованием Windows Forms [4], который предоставляет простой и интуитивно понятный способ создания графических пользовательских интерфейсов (GUI) для приложений, используя дизайнер форм и компоненты управления.

Несмотря на другие решения для реализации поставленной задачи, в том числе макросы, было решено разработать именно полноценную программу. Во-первых, подразумевается, что параметризируются сложные технические устройства, в нашем случае, принтер, которые содержат больше сотни различных больших и мелких деталей, каждая из которых имеет свою координату в пространстве, а некоторые изменяемую длину и ширину. Изменения одного из основных размеров принтера могут привести к изменениям чуть ли не во всей сборке. Во-вторых, API Компас-3D предоставляет большой выбор для программного изменения модели сборки, например, материал детали. Таким образом, реализация программы даёт больше возможностей для развития в будущем.

Заключение

Актуальность 3D-печать становится все более очевидной в современном мире, предлагая революционные возможности в различных отраслях, от промышленного производства до медицинской сферы и дизайна. Важной особенностью технологии является возможность создания регулируемых, индивидуальных продуктов, отвечающих уникальным потребностям конечного пользователя. Это открывает новые горизонты в области персонализации и инноваций, позволяя производителям и дизайнерам

создавать уникальные изделия, которые ранее были невозможны или экономически нецелесообразны для производства.

Список литературы

1. Официальный сайт программы КОМПАС-3D, URL: <https://kompas.ru/>
2. Герасимов, А. А. Самоучитель "КОМПАС-3D" V19. Санкт-Петербург: БХВ-Петербург, 2021. 624 с., ил. (Самоучитель).
3. Документация по C#. Начало работы, руководства, справочные материалы | Microsoft Learn [Электронный ресурс].
4. Обзор - Windows Forms .NET Framework | Microsoft Docs [Электронный ресурс]. URL: <https://docs.microsoft.com/ru-ru/dotnet/desktop/winforms/windows-formsoverview?view=netframework-desktop-4.8>

СТРУКТУРА, ПРОВОДИМОСТЬ И СЕНСОРНЫЕ СВОЙСТВА ГИДРОТЕРМАЛЬНЫХ КОМПОЗИТОВ ZnO-In₂O₃

Ерофеева А.Р.^{1,2}, Иким М.И.²

¹ *Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ,*

² *Федеральный исследовательский центр
химической физики им. Н.Н. Семёнова
Российской академии наук*

Аннотация

Исследовано влияние состава гидротермальных композитов ZnO-In₂O₃ на структуру, проводимость и сенсорные свойства. Введение до 20 мас.% ZnO в композит приводит к снижению сенсорного отклика на водород, а добавка 65 мас.% ZnO увеличивает отклик почти в два раза по сравнению с чистым In₂O₃. На сенсорную активность влияет фазовый состав In₂O₃, который изменяется в зависимости от содержания ZnO. Максимальная эффективность достигается для композита, содержащего кубический In₂O₃, и минимальная для ромбоэдрического In₂O₃.

Введение

В связи с увеличением числа промышленных предприятий, деятельность которых негативно влияет на экологическую обстановку, возникает потребность в обнаружении в атмосфере взрыво- и пожароопасных соединений, а также вредных для здоровья газов. Ключевое значение приобретает создание эффективных и избирательных сенсоров для определения различных веществ в атмосфере. Наиболее перспективные в этом плане полупроводниковые сенсоры, которые отличаются доступной стоимостью, простотой производства и обслуживания.

Датчики, основанные только на одном оксиде металла, обладают недостаточно высокой эффективно-

стью и селективностью при обнаружении многокомпонентных газовых смесей. Установлено, что для получения эффективных сенсоров системы должны содержать смесь оксидов, один из которых имеет высокую концентрацию электронов проводимости, а другой – высокую каталитическую активность [1].

Одним из оксидов металлов, широко используемых для изготовления сенсоров, является In_2O_3 – полупроводник n-типа с высокой концентрацией электронов проводимости и шириной запрещенной зоны 2,8 эВ [2].

Эффективность сенсора на основе In_2O_3 может быть значительно повышена добавлением оксида цинка к чувствительному слою [3-5]. ZnO имеет широкую запрещенную зону 3,2 эВ и отличную химическую и термическую стабильность.

В настоящей работе гидротермальным методом синтезированы композиты $\text{ZnO-In}_2\text{O}_3$, содержащие 5-85 % оксида цинка, а также оценено влияние концентрации ZnO в композитах на их фазовый состав, структуру, проводимость и сенсорные свойства при детектировании водорода.

Экспериментальная часть

Композиты $\text{ZnO-In}_2\text{O}_3$, содержащие от 5 до 85 мас.% оксида цинка, были получены гидротермальным методом.

Состав, морфологические и структурные параметры синтезированных систем исследовали с помощью различных методов, таких как рентгеновская дифракция, сканирующая электронная микроскопия, энергодисперсионная рентгеновская спектроскопия и низкотемпературная газовая адсорбция.

Для определения проводимости и сенсорных свойств синтезированные композиты формировали в чувствительный слой. Сенсорный отклик полученных пленок на H_2 исследовали с помощью разработанной установки в диапазоне температур от 280 до 520 °С. Точность поддержания температуры находилась в пределах 1 °С. Чип с нанесенным чувствительным слоем располагался в специальной камере объемом около 1 см³, в которую подавался очищенный воздух или калибровочная газовая смесь, содержащая 0,1% H_2 . Газы прокачивались через камеру со скоростью 200 мл/мин.

Чувствительность S (сенсорный отклик) определяли, как величину отношения $S = R_o/R$, где R_o – начальное сопротивление сенсора (до введения анализируемой смеси), R – минимальное значение сопротивления сенсора, достигаемое после введения анализируемого газа.

Структурные характеристики композитов $\text{ZnO-In}_2\text{O}_3$

На дифрактограммах композитов $\text{ZnO-In}_2\text{O}_3$ зарегистрированы пики, относящиеся как к кубической, так и к ромбоэдрической фазам In_2O_3 . В XRD спектрах гидротермальных образцов, содержащих до 20% ZnO , отсутствуют дифракционные пики, характерные для ZnO . Это можно объяснить тем, что ионы Zn встраиваются в кристаллическую решётку оксида

индия либо формируется рентгеноаморфная оксидная фаза.

Ромбоэдрическая фаза In_2O_3 фиксируется в спектрах XRD только при концентрациях от 10 до 40% ZnO в композитах. Образование ромбоэдрической фазы In_2O_3 происходит за счет частичного растворения ионов Zn , имеющих меньший размер, чем ионы In , в решетке основных кубических кристаллов In_2O_3 . Отметим, что концентрация ромбоэдрической фазы оксида индия максимальна в композитах, содержащих 20% ZnO . В композитах с концентрацией больше 40% ZnO наблюдается только кубическая фаза оксида индия.

На дифрактограммах композитов $\text{ZnO-In}_2\text{O}_3$, содержащих от 20 до 85% ZnO , появляются пики, характерные ZnO со структурой вюрцита (JCPDS № 36-1451). Отсутствие посторонних пиков свидетельствует о высокой чистоте и кристалличности образцов.

По данным РФА, увеличение концентрации ZnO в композитах сопровождается уменьшением размера наночастиц оксида индия с 25 до 9 нм. Этот результат свидетельствует о том, что в процессе синтеза атомы Zn замещают атомы In в решетке, что создает деформации в композитах. Таким образом, введение ионов Zn в структуру In_2O_3 препятствует росту кристаллов. В композитах, содержащих от 20 до 85 мас.% оксида цинка, размер частиц ZnO почти в три раза больше, чем у In_2O_3 , и уменьшается с 72 до 25 нм. Наблюдаемая разница в размерах наночастиц оксида цинка и оксида индия может свидетельствовать о том, что оксид цинка полностью покрывает поверхность оксида индия.

Согласно данным SEM и EDX композит 5% $\text{ZnO-95\%In}_2\text{O}_3$ состоит из частиц оксида индия с неоднородной, рыхлой структурой, в которой ионы цинка распределены достаточно равномерно.

Изотермы адсорбции-десорбции гидротермального композита имеют форму IV типа с петлей гистерезиса H3. Такая форма характерна для мезопористой природы образцов. Площадь поверхности, рассчитанная методом БЭТ, составила 25 м²/г. Для изучения мезопористой структуры использовали метод ВЖН для десорбционной ветви изотермы. Размер пор образцов составляет 18–20 нм. Следует отметить, что введение оксида цинка приводит к увеличению удельной поверхности образца, объема и размера пор.

Проводимость и сенсорные свойства композитов $\text{ZnO-In}_2\text{O}_3$

В диапазоне от 300 до 520 °С проводимость всех синтезированных систем увеличивается с ростом температуры.

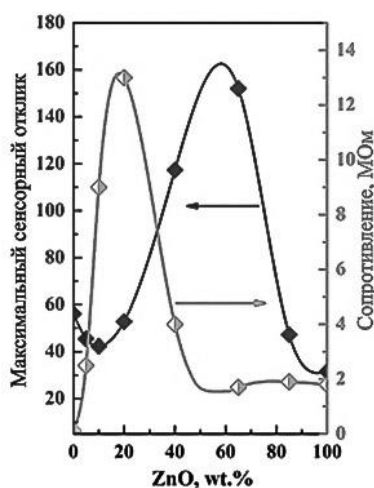


Рис. 1. Влияние концентрации ZnO на максимальный отклик сенсора при 1100 ppm H₂ и сопротивление образцов

В системах, содержащих до 20% ZnO, ионы цинка внедряются в кристаллическую структуру оксида индия. При этом электронная структура оксида индия изменяется, а увеличение сопротивления In₂O₃ происходит за счет уменьшения концентрации электронов проводимости в этом оксиде (рис. 1). Образование «дырок» в системе при замещении ионов индия в решетке In₂O₃ на ионы цинка, которые являются акцепторами, влияет на рост сопротивления в композитах, содержащих до 20% оксида цинка. В этом случае электроны компенсируются дырками, генерируемыми акцепторами, и, как следствие, сопротивление увеличивается.

В структуре композитов, содержащих более 20% ZnO, появляются кристаллы оксида цинка, которые при взаимодействии с модифицированными кристаллами In₂O₃ приводят к резкому снижению сопротивления композита.

На кривой зависимости максимального сенсорного отклика от состава можно выделить несколько характерных участков: 0–20 мас.%, 20–85 мас.% и выше 85 мас.% ZnO (рис. 1). Эти участки различаются по влиянию оксида цинка на сенсорный отклик композита. Реакция сенсора на водород снижается в композитах, содержащих до 20 мас.% ZnO, по сравнению с чистым оксидом индия. По данным рентгенофазового анализа, в таких композитах не образуется кристаллический оксид цинка, а ионы цинка внедряются в структуру оксида индия.

На величину сенсорного отклика гидротермальных композитов ZnO-In₂O₃, содержащих до 20 % оксида цинка, существенное влияние оказывает тот факт, что в процессе синтеза таких композитов происходит изменение фазового состава In₂O₃, а именно — переход кубической фазы в ромбоэдрическую.

Увеличение концентрации оксида цинка с 20 до 85 мас.% приводит к увеличению сенсорного отклика на водород почти в два раза. При этом уже происходит образование кристаллических частиц оксида цинка и возвращение стабильной кубической фазы оксида индия. Увеличение сенсорного отклика на

водород в композитах, содержащих от 20 до 85% оксида цинка, обусловлено, с одной стороны, образованием гетероперехода между ZnO и In₂O₃ и, с другой стороны, уменьшением доли ромбоэдрической фазы оксида индия.

Отметим, что максимальной сенсорной активностью обладает композит, содержащий 65% ZnO. Согласно полученным данным, в такой системе образуются только две фазы (кубический оксид индия и гексагональный оксид цинка), а усиление сенсорной реакции связано с химической сенсibilizацией. Снижение сенсорного отклика композитов, содержащих более 85% оксида цинка, связано с прерыванием протекания тока между оксидами индия и цинка. Это связано с тем, что наночастицы кристаллического ZnO образуются на In₂O₃ при гидротермальном синтезе и при такой концентрации полностью покрывают поверхность оксида индия. Соответственно, чувствительность нанокompозита снижается, приближаясь к чувствительности нанокристаллического ZnO.

Заключение

Проведённое исследование показало, что фазовый состав, структура, проводимость и сенсорные свойства синтезированных систем сильно зависят от концентрации оксида цинка.

В композитах, содержащих до 40 мас.% ZnO, взаимодействие наночастиц приводит к изменению фазового состава и структуры частиц In₂O₃ в результате внедрения ионов цинка в In₂O₃. При этом наблюдается уменьшение сенсорного отклика на водород.

Напротив, максимальный сенсорный отклик в таких системах обнаружен при концентрации 65% ZnO, когда происходит формирование двухфазной системы и возвращение оксида индия в стабильную кубическую фазу.

Список литературы

- Gerasimov G. N. et al. The mechanisms of sensor phenomena in binary metal-oxide nanocomposites //Sensors and Actuators B: Chemical. – 2017. – Т. 240. – С. 613-624.
- Wang Z. et al. Highly photocatalytic ZnO/In₂O₃ heteronanostructures synthesized by a coprecipitation method //The Journal of Physical Chemistry C. – 2009. – Т. 113. – №. 11. – С. 4612-4617.
- Zhu L., Li Y., Zeng W. Hydrothermal synthesis of hierarchical flower-like ZnO nanostructure and its enhanced ethanol gas-sensing properties //Applied Surface Science. – 2018. – Т. 427. – С. 281-287.
- Jiang B. et al. Construction of mesoporous In₂O₃-ZnO hierarchical structure gas sensor for ethanol detection //Sensors and Actuators B: Chemical. – 2023. – С. 134203.
- Kang Y. et al. Review of ZnO-based nanomaterials in gas sensors //Solid State Ionics. – 2021. – Т. 360. – С. 115544.

Секция

«Низкоразмерные квантовые системы»

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ НЕРАВНОВЕСНОГО КВАЗИЧАСТИЧНОГО ТРАНСПОРТА В СВЕРХПРОВОДНИКЕ

Гурский А.С.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В настоящей работе впервые обнаружена когерентная составляющая электронного транспорта неравновесных квазичастиц в экспериментальных образцах гибридных структур металл-изолятор-сверхпроводник в виде твердотельного интерферометра. Получены экспериментальные данные зависимости туннельного тока от внешнего магнитного поля на твердотельном интерферометре, качественно подтверждающие искомым эффект.

Введение

Одним из потенциальных решений для проблем высокого тепловыделения в классической электронике может стать использование сверхпроводящих элементов, отличающихся пренебрежимо малым тепловыделением и технологической доступностью в изготовлении. Однако при проектировании устройств на сверхпроводниковой базе должны учитываться явления, возникающих на границах (интерфейсах) различных материалов, для повышения стабильности и, соответственно, времени работы. Учет таких эффектов становится принципиально важным, когда размеры элементов сравниваются с характерными длинами релаксации. Для изучения физики обозначенной проблемы предлагается экспериментально исследовать возможность наблюдения когерентного транспорта неспаренных электронов в гибридных сверхпроводящих структурах металл-изолятор-сверхпроводник на базе твердотельного интерферометра. Необходимость изучения подобных процессов, обуславливающих работу широкого класса нанoelectronных устройств – например, сверхпроводниковых кубитов [1-4], делает предлагаемые к исследованию в настоящей работе задачи крайне актуальными и предлагает результаты, имеющие прямое отношение к работе многих квантовых устройств [5-7].

Релаксация квазичастиц в сверхпроводнике

Известно, что в спектре элементарных возбуждений сверхпроводящего состояния, в отличие от нормального металла, существует энергетическая щель 2Δ , эквивалентная энергии связи куперовской пары. Из-за термической активации в спектре сверхпроводника всегда представлено некоторое число неспаренных (нормальных) электронов, соответствующих так называемым *равновесным* квазичастичным возбуждениям.

Если на систему оказывается дополнительное воздействие, приводящее к разрушению куперовских пар, то также появляются дополнительные, *неравновесные*, квазичастицы.

Для нормального металла в пределе сверхнизких температур все состояния ниже уровня Ферми E_F будут заняты, а выше – свободны. Основное отличие сверхпроводящего состояния в таком случае заключается в наличии энергетической щели 2Δ . Для удобства спектр обычно смещают вверх на E_F – тогда уровень Ферми соответствует нулю энергии, и все занятые состояния обладают отрицательной энергией, а незанятые – положительной (рис. 1). При этом состояния с импульсом (и энергией E_k), меньшим по модулю фермиевского ($|p| < |p_F|$), называются дырочными, а больше – электронными ($|p| > |p_F|$). При разрушении куперовских пар возникает пара неспаренных электронов с противоположными по знаку импульсами.

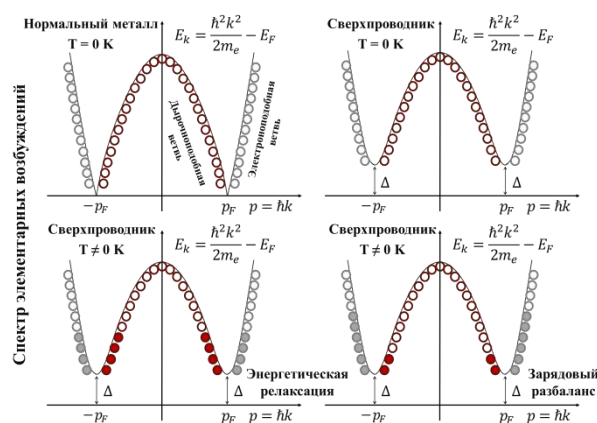


Рис. 1. Спектры возбуждений обычного металла и сверхпроводника при нулевой и конечной температуре. Электроноподобные возбуждения показаны серым цветом, а дырочноподобные – красным

При нуле температур все носители заряда представлены парными электронами, куперовскими парами, с нулевым суммарным импульсом. При больших температурах, появляются равновесные квазичастичные возбуждения, и после разрушения куперовских пар в спектре возбуждений равным образом заселяются обе ветви – это случай *продольной*, или *энергетической моды* возбуждений.

Возможно и нарушение симметрии заселения спектра, когда в сверхпроводник (S) из нормального металла (N) инжектируются преимущественно электроны или дырки – это так называемая *поперечная мода*, или *зарядовый разбаланс*. В таком случае, при одной полярности электрического напряжения, приложенного на N/S контакт, можно преимущественно заселять электроноподобные или, при обратной – дырочноподобные ветви.

Из-за неравномерного заселения в спектре возникает несколько процессов. Первый, и самый быстрый из них, соответствует рассеянию квазичастиц на времени τ_s порядка нескольких фемпторосекунд – на меньших временах бессмысленно говорить о температуре квазичастиц. На выравнивание перекоса в спектре элементарных возбуждений сверхпроводника требуется характерное время τ_Q . Наконец, за время τ_R , из неравновесных квазичастиц образуется равновесная куперовская пара. Главным образом, описанные процессы являются неупругими и, соответственно, требующими некоторую дополнительную систему для обмена энергией – чаще всего, такую роль выполняют колебания решетки, фононы. Однако в области температур $T \ll T_c$ подобный обмен затруднителен, из-за чего неравновесность спектра может существовать довольно продолжительное время и на значительных для мезоскопичеких длинах.

При сверхнизких температурах длины квазичастичной релаксации часто оказываются намного больше характерных размеров реальных сверхпроводящих гибридных структур. В качестве примера можно привести сверхпроводящий алюминий (Al), у которого длина зарядовой релаксации λ_Q составляет около 5 μm , а энергетическая λ_E – имеет величину порядка 40 μm [8, 9]. Для сравнения стоит отметить, что длина свободного пробега электрона l около 30 нм и сверхпроводящая длина когерентности $\xi_S \sim 100$ нм. Таким образом, структуры (суб)микронного размера, изготовленные на базе алюминия, существуют в существенно неравновесном состоянии.

Экспериментальные методы и образцы

С использованием метода углового вакуумного напыления, были изготовлены гибридные наноструктуры, представляющие сверхпроводящую алюминиевую «вилку» толщиной 30 нм, отделенную через тонкий слой изолятора (AlOx) от нормального медного электрода, толщиной 50 нм. Данные структуры можно формально рассматривать в качестве твердотельных аналогов классического оптического интерферометра (рис. 2).

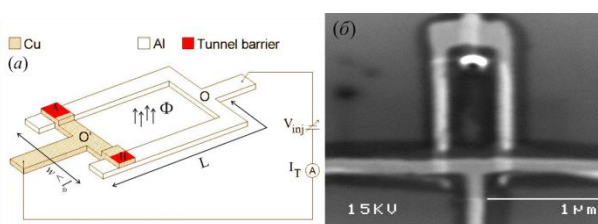


Рис.2. (а) Схематичное представление твердотельного NIS интерферометра. (б) Микрофотография NIS интерферометра, полученная методом растровой электронной микроскопии

Ширина нормального электрода была выбрана меньше характерной длины сбега фазы в меди $l_\varphi \sim 1$ μm , что соответствует инжекции одного и того же неравновесного носителя заряда на оба туннельных

контакта. Исходя из предложенной гипотезы, должна существовать некоторая характерная длина λ_φ , на которой сохраняется фаза волновой функции неспаренных частиц в сверхпроводнике, поэтому при длинах плеч интерферометра меньше ее ($L < \lambda_\varphi$), станет возможным наблюдение интерференции – когерентной составляющей тока квазичастиц.

Для достижения необходимой области сверхнизких температур порядка 10 мК использовался рефрижератор растворения Kelvinox 400 НА с множественными защитными экранами и фильтрами [10], позволяющими минимизировать влияние внешнего излучения на образцы. Все измерения были произведены на постоянном токе и/или с использованием фазочувствительных детекторов.

Результаты

Типичные для контакта металл-изолятор-сверхпроводник вольт-амперные характеристики (ВАХ) показаны на рис. 3. В частности, излом ВАХ соответствует энергетической щели алюминия $V_{gap} = \Delta/2e \approx 210$ мкВ. При фиксации энергии инжекции выше щели $eV = const \geq \Delta$ туннельный ток I_T через контакт металл-изолятор-сверхпроводник измерялся как функция перпендикулярного магнитного поля.

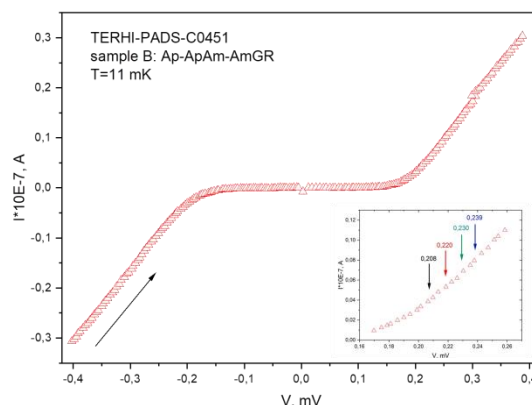


Рис.3. Вольт-амперная характеристика $I(V)$ NIS интерферометра, полученная при температуре $T = 11$ мК. На вставке показан фрагмент ВАХ вблизи щели Δ/e со значениями напряжений смещения V_{bias} , при которых производилась запись зависимостей $I_T(V = V_{bias}, B)$

В случае движения по контуру $O'A'O'B'O'$, у квазичастиц появляется набег фазы $\delta\varphi$, приводящий к наблюдению интерференции. Разность фаз в разных плечах интерферометра достигается за счет эффекта Ааронова-Бома, приводящего к периодической модуляции фазы волновой функции неравновесной квазичастицы с периодом Φ/Φ_0 , где квант магнитного потока $\Phi_0 = h/(2e) \approx 2.07 \times 10^{-15}$ Вб. Наблюдение немонойтонной зависимости для туннельного тока в этом случае и будет ассоциироваться с проявлением когерентного транспорта неравновесных квазичастиц.

После приложения на NIS контакт смещающего электрического напряжения V_{bias} выше щелевого,

измерялся туннельный ток I_T при изменении магнитного поля. Все зависимости $I_T(V = V_{bias}, B)$ на рис. 4 имеют выраженный «центральный максимум» в диапазоне от -1 мТл до 1 мТл и некоторые немоности, которые могут быть вторичными максимумами. Принимая во внимание площадь интерферометра, номинально равную $0.4 \text{ мкм} \times 1.2 \text{ мкм} = 0.48 \text{ мкм}^2$, можно высчитать характерный период осцилляций $\delta B \approx 4.1 \text{ мТл}$, что с высокой точностью соответствует положению неоднородностей, представленных на рис. 4.

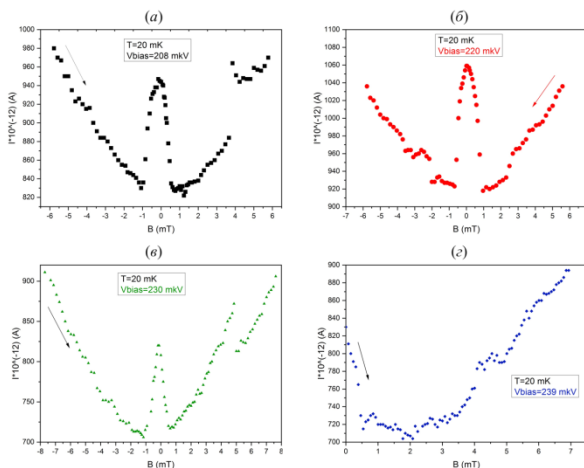


Рис. 4. Зависимости туннельного тока от магнитного поля $I_T(V = V_{bias}, B)$ для NIS интерферометра, измеренные при температуре $T = 20 \text{ мК}$ и напряжениях смещения $V_{bias} = 208, 220, 230 \text{ и } 239 \text{ мКВ}$

Результаты, полученные для туннельного тока, качественно согласуются с оптической интерференционной картиной для классического двухлучевого интерферометра. Относительно монотонный рост туннельного тока I_T при увеличении магнитного поля и заданном смещении V_{bias} может объясняться частичным подавлением сверхпроводящего состояния. Предполагается, что именно эта особенность не позволяет разрешить максимумы более высоких порядков.

В связи с вышесказанным, можно заключить, что в данной работе действительно удалось впервые обнаружить крайне интересные зависимости, отражающие нетривиальную неравновесную физику мезоскопических систем и требующие дальнейшего изучения.

Заключение

Проведенное исследование экспериментально доказывает возможность наблюдения когерентной составляющей квазичастичного тока в гибридных структурах типа металл-изолятор-сверхпроводник в виде твердотельного интерферометра. Было продемонстрировано, что туннельный ток немонотонным образом зависит от магнитного поля при постоянном напряжении инжекции. В дальнейшем планируется исследовать новое поколение образцов с различной длиной плеч интерферометра. С одной стороны, большая площадь интерферометров должна соответ-

ствовать меньшему периоду интерференционной картины в магнитном поле, что приводит к более слабой монотонной составляющей в зависимостях $I_T(V = V_{bias}, B)$ и потенциально может повысить разрешающую способность максимумов более высокого порядка. С другой стороны, увеличение длины плеч интерферометра L в конечном итоге приведет к исчезновению эффекта $L > \lambda_\phi$, что позволит оценить величину λ_ϕ .

Список литературы

1. Measurement of energy decay in superconducting qubits from nonequilibrium quasiparticles / M. Lenander, H. Wang, R. C. Bialczak [et al.] // *Physical Review B: Condensed Matter and Materials Physics*. – 2011. – Vol. 84, No. 2. – P. 024501.
2. Decoherence of superconducting qubits caused by quasiparticle tunneling / G. Catelani, S. E. Nigg, S. M. Girvin [et al.] // *Physical Review B: Condensed Matter and Materials Physics*. – 2012. – Vol. 86, No. 18. – P. 184514.
3. Zanker, S. Qubit dephasing due to quasiparticle tunneling / S. Zanker, M. Marthaler // *Physical Review B: Condensed Matter and Materials Physics*. – 2015. – Vol. 91, No. 17. – P. 174504.
4. Hot Nonequilibrium Quasiparticles in Transmon Qubits / K. Serniak, M. Hays, G. De Lange [et al.] // *Physical Review Letters*. – 2018. – Vol. 121, No. 15. – P. 157701.
5. Effective electron microrefrigeration by superconductor-insulator-normal metal tunnel junctions with advanced geometry of electrodes and normal metal traps / I. J. Agulo, L. Kuzmin, M. Fominisky, M. Tarasov // *Nanotechnology*. – 2004. – Vol. 15, No. 4. – P. S224-S228.
6. Cooling electrons from 1 to 0.4 K with V-based nanorefrigerators / O. Quaranta, P. Spathis, F. Beltram, F. Giazotto // *Applied Physics Letters*. – 2011. – Vol. 98, No. 3. – P. 032501-3.
7. Semenov, A. D. Hot-electron effect in superconductors and its applications for radiation sensors / A. D. Semenov, G. N. Gol'tsman, R. Sobolewski // *Superconductor Science and Technology*. – 2002. – Vol. 15, No. 4. – P. R1-R16.
8. Arutyunov, K. Y. Spatially resolved measurement of nonequilibrium quasiparticle relaxation in superconducting Al / K. Y. Arutyunov, H. P. Auraneva, A. S. Vasenko // *Physical Review B: Condensed Matter and Materials Physics*. – 2011. – Vol. 83, No. 10. – P. 104509.
9. Nonlocal supercurrent in mesoscopic multiterminal SNS Josephson junction in the low-temperature limit / T. E. Golikova, I. E. Batov, I. V. Bobkova [et al.] // *Physical Review B: Condensed Matter and Materials Physics*. – 2014. – Vol. 89, No. 10. – P. 104507.
10. Examination of cryogenic filters for multistage RF filtering in ultralow temperature experiments / V. V. Zavyalov, K. Y. Arutyunov, S. A. Chernyaev [et al.] // *Journal of Physics: Conference Series*: 28, Gothenburg, 09–16 августа 2017 года. – Gothenburg, 2018. – P. 012086.

РАЗМЕРНЫЕ ЭФФЕКТЫ В СУБМИКРОННЫХ ПЛЕНКАХ ПОЛИАРИЛЕНФТАЛИДОВ

*Галиев А.Ф., Карамов Д.Д., Каримов В.Р.,
Ишмухаметов М.С., Буланкин Н.С.
Башкирский государственный педагогический
университет им. М.Акумлы,
кафедра физики и нанотехнологий
Института физики, математики,
цифровых и нанотехнологий*

Аннотация

Наряду с активными исследованиями полупроводниковых и металлоподобных органических материалов, таких как электропроводящие полимеры, интерес представляют эффекты аномального переноса заряда в тонких слоях пленок оксидов металлов, а также органических диэлектриков. В тонких пленках различных полимеров, таких как полипропилен, полиимид и др., не превышающих десятков микрон, наблюдаются явления, связанные с фазовым переходом из диэлектрического состояния в высокопроводящее.

Было обнаружено, что такой переход может быть инициирован внешними физическими воздействиями: механическим давлением, температурой, магнитными и электрическими полями и другими. Несмотря на то, что механизмы проводимости полимерных пленок достаточно хорошо изучены, состояниям с высокой проводимостью уделяется недостаточно внимания. В частности, остается неясной связь между химической структурой полимера, супрамолекулярной организацией и электрическими свойствами пленок; влияние барьерных эффектов на границе раздела металл/полимер на процессы переноса заряда; характеристики высокопроводящего состояния. Кроме того, регистрация изменений работы выхода электрона из металла вследствие фазовых переходов, а также возникновения упругих и неупругих деформаций открывает новое направление диагностики состояния металлов.

Как известно, большинство немодифицированных полимерных материалов по своей природе являются диэлектриками [1, 2]. На основе зонной теории твердого тела известно, что диэлектрики идеальной структуры не должны проводить электрический ток. На локальных участках материала это можно наблюдать, и полимеры не являются исключением, то есть в полимерных пленках всегда имеются структурные нарушения, дефекты и примеси, вызывающие генерацию подвижных носителей заряда.

Одним из интересных электронных свойств органических материалов является способность менять состояние с диэлектрика или полупроводника на состояние с высокой проводимостью [3]. Это явление называется эффектом переключения проводимости или просто эффектом переключения.

Эффекты переключения и памяти

Эффект переключения проявляется следующим образом. В обычном состоянии при приложении

электрического поля к тонкой пленке ток через образец возрастает линейно. Ток в этом состоянии мал, то есть сопротивление образца велико (по закону Ома для участка цепи $I=U/R$). При достижении определенного значения напряжения $U_{\text{выкл}}$, называемого пороговым или критическим, наблюдается резкое увеличение величины электрического тока через образец.

При движении в сторону уменьшения напряжения коммутируемое состояние будет сохраняться до определенного значения напряжения $U_{\text{выкл}}$, после чего образец возвращается в исходное «выключенное» состояние. При этом эффекте зависимость тока от напряжения, то есть вольт-амперная характеристика (ВАХ), имеет S-образный вид. Примечательно, что изменение сопротивления может достигать нескольких порядков, то есть наблюдается переход материала из диэлектрика или полупроводника в высокопроводящее состояние.

На рис.1 представлена модель локального переключения полимерной пленки из диэлектрического состояния в состояние с высокой проводимостью.

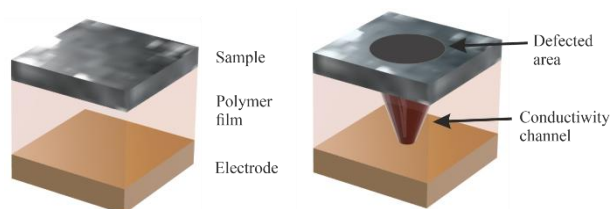


Рис.1. Эффект переключения полимерной пленки из диэлектрического в высокопроводящее состояние при деформации металла

Также можно наблюдать второй тип переключения, называемый эффектом памяти. Отличие этого эффекта от переключения состоит в том, что состояние высокой проводимости сохраняется даже после выключения напряжения на образце, а при последующем приложении напряжения образец находится в состоянии высокой проводимости [3].

Изменение электропроводности органического материала может быть спровоцировано незначительными внешними факторами. Одним из таких факторов может быть изменение электронных параметров в металлическом контакте при избыточном механическом давлении.

Результаты представлены на рис.2. Видно, что пороговое давление, при котором наблюдается эффект переключения [15], зависит от типа металла, а следовательно, и его электронных характеристик.

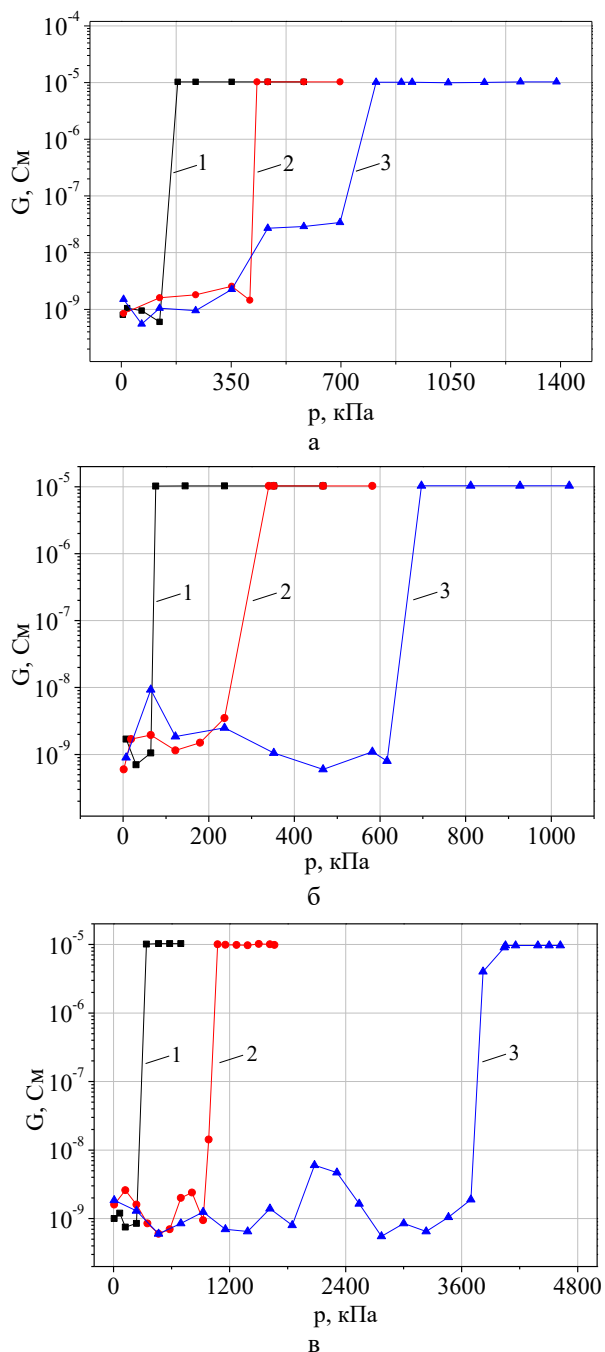


Рис.2. Зависимости проводимости от приложенного одноосного механического давления для образцов, имеющих разную толщину полимерной пленки (1 ~ 40 нм, 2 ~ 280 нм, 3 ~ 1300 нм): а) материал электрода – медь; б) материал электрода – алюминий; в) материал электрода – хром

В то же время, на графиках наблюдается нетривиальная зависимость порогового давления от толщины полимерной пленки. Вопреки ожидаемой линейной зависимости критического давления от толщины, видно, что при толщинах менее 100 нм, критические давления возрастают. По-видимому, это связано с особенностями переноса заряда в тонких пленках, а равно с особенностями структуры на надмолекулярном уровне организации.

Исходя из полученных данных, можно сделать вывод, что эффекты переключения в полимерных пленках могут быть использованы при разработке устройств памяти или неразрушающего контроля.

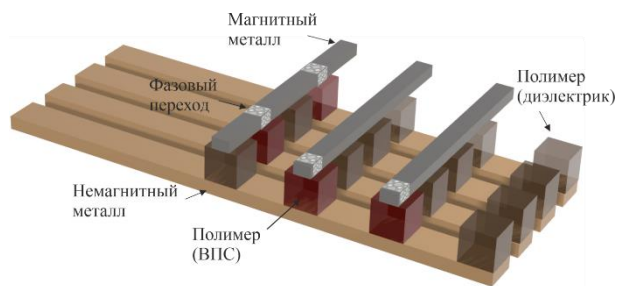


Рис.3. Схематическое изображение матричного сенсора механического напряжения в металлическом образце

На рис.3 представлена возможная схема реализации такого устройства.

Заключение

Изменение работы выхода электронов металла - факт, свидетельствующий о различных процессах в металлах. Это могут быть химические изменения, в частности окисление поверхностных слоев. При этом практический интерес представляют изменения, связанные со структурой – возникновение трещин, усталость металла, упругие и неупругие деформации. Отслеживание изменений работы выхода может быть методом определения всех этих преобразований.

В работе использовался электроактивный полимер. Вероятно, изменяя химический состав полимера, а также его толщину, можно добиться селективности к определенным видам структурных изменений металлов.

Благодарность

Работа выполнена в рамках программы «Зеркальные лаборатории» Национального исследовательского университета Высшая школа экономики и Башкирского государственного педагогического университета им. М.Акумуллы.

Список литературы

- Zhou, L., Zhao, S., Xie, P., Miao, X., Liu, S., Sun, N., ... & Shen, Y. (2023). Research progress and prospect of polymer dielectrics. *Applied Physics Reviews*, 10(3). <https://doi.org/10.1063/5.0151215>
- Shanmugasundram, H. P. P. V., Jayamani, E., & Soon, K. H. (2022). A comprehensive review on dielectric composites: Classification of dielectric composites. *Renewable and Sustainable Energy Reviews*, 157, 112075. <https://doi.org/10.1016/j.rser.2022.112075>
- Paul, F., & Paul, S. (2022). To be or not to be – review of electrical bistability mechanisms in polymer memory devices. *Small*, 18(21), 2106442. <https://doi.org/10.1002/sml.202106442>

РЕШЕНИЕ УРАВНЕНИЙ УЗАДЕЛЯ ДЛЯ КОНТАКТА ФЕРРОМАГНИТНЫЙ ИЗОЛЯТОР-СВЕРХПРОВОДНИК-ФЕРРОМАГНИТНЫЙ ИЗОЛЯТОР В ЛИНЕЙНОМ ПРИБЛИЖЕНИИ

Сеидов С.С., Селезнёв Д.В., Пугач Н.Г.

Национальный исследовательский университет

«Высшая школа экономики»,
научно-учебная лаборатория
квантовой наноэлектроники
МИЭМ НИУ ВШЭ

Аннотация

В работе рассматривается контакт ферромагнитный изолятор-сверхпроводник-ферромагнитный изолятор в грязном пределе. В линейном приближении для функций Грина электронов в сверхпроводнике записаны и решены уравнения Узаделя с соответствующими граничными условиями. Найдена энергия контакта при параллельной и антипараллельной ориентациях ферромагнитных изоляторов с обеих сторон от сверхпроводника.

Введение

При контакте сверхпроводника с несверхпроводящим материалом на границе соприкосновения возникает так называемый эффект близости. Если материал является проводником, то в нём благодаря туннелированию куперовских пар из сверхпроводника в тонком слое вблизи границы возникает наведённая сверхпроводимость [1]. Если же материал является тонким изолятором, то при расположении с другой стороны второго сверхпроводника, возможно создание слабой связи между двумя сверхпроводниками – джозефсоновского контакта. Особый интерес представляют магнитные несверхпроводящие материалы. Так как их магнитный момент взаимодействует с собственным магнитным моментом входящих в куперовскую пару электронов, пространственная ориентация магнетика влияет на свойства гетероструктуры. При этом через непроводящий магнетик возможно прохождение спинового тока в зависимости от его ориентации. Данное свойство открывает возможности для применения гетероструктур сверхпроводник-магнитный изолятор в сверхпроводящей спинтроники, где магнетик будет играть роль затвора для спинового тока через гетероструктуру.

Теоретическое описание контакта

Рассмотрим сверхпроводник, с двух сторон обложенный ферромагнитными изоляторами, как это изображено на рис. 1.



Рис. 1. Гетероструктура ферромагнитный изолятор-сверхпроводник-ферромагнитный изолятор

Он описывается матричной функцией Грина

$$\hat{g} = \begin{pmatrix} g_{\uparrow\uparrow} & 0 & f_{\uparrow\uparrow} & f_{\uparrow\downarrow} \\ 0 & g_{\uparrow\uparrow} & f_{\downarrow\uparrow} & f_{\downarrow\downarrow} \\ -f_{\uparrow\uparrow} & -f_{\downarrow\uparrow} & -g_{\uparrow\uparrow} & 0 \\ -f_{\downarrow\uparrow} & -f_{\downarrow\downarrow} & 0 & -g_{\uparrow\uparrow} \end{pmatrix}.$$

Предполагается, что сверхпроводник s-типа, поэтому антипараллельные функции Грина $g_{\uparrow\downarrow}$ и $g_{\downarrow\uparrow}$ положены равными нулю. Однако все аномальные компоненты сохранены, так как они возникнут благодаря взаимодействию с ферромагнитными изоляторами.

В грязном пределе движение электронов в гетероструктуре диффузионное и описывается уравнением Узаделя

$$iD\partial_z(\hat{g}\partial_z\hat{g}) = [\varepsilon\tau_3 \otimes \sigma_0 + \hat{\Delta}, \hat{g}].$$

Здесь D – константа диффузии, ε – энергия квазичастиц, σ и τ – матрицы Паули в пространстве спина и Намбу соответственно, $\hat{\Delta} = \text{antidiag}(+\Delta, -\Delta, +\Delta^*, -\Delta^*)$ – сверхпроводящий параметр порядка.

Энергия контакта даётся функционалом [2, 3]

$$F = \frac{|\Delta|^2}{2} + \int dz d\varepsilon \text{Tr} \left[\frac{i}{4} D(\partial_z \hat{g})^2 + (\varepsilon\tau_3 \otimes \sigma_0 - \hat{\Delta}) \hat{g} \right].$$

Его же вариация приводит к уравнениям Узаделя. Граничные условия для уравнения Узаделя имеют вид [4]

$$G_L L_L(\hat{g}_L \partial_z \hat{g}_L) = -G_R L_R(\hat{g}_R \partial_z \hat{g}_R) = 2I.$$

Индексы L и R соответствуют левой и правой границе. Матрица I называется матричным током и для ферромагнитного изолятора имеет вид

$$I = -NG_Q \left[1 - \frac{i}{4} \sin(\varphi) \hat{a} + \frac{1}{2} \sin^2\left(\frac{\varphi}{2}\right) \hat{a}\hat{m} \right]^{-1} \times \left[-i \sin(\varphi) \hat{g}\hat{a} + \sin^2\left(\frac{\varphi}{2}\right) [\hat{m}, \hat{a}] \right] \times \left[1 - \frac{i}{4} \sin(\varphi) \hat{a} + \frac{1}{2} \sin^2\left(\frac{\varphi}{2}\right) \hat{m}\hat{a} \right]^{-1}.$$

Здесь φ – угол спинового смешивания, $\hat{a} = \hat{g}\hat{m}\hat{g} - \hat{m}$. Матрица \hat{m} описывает ориентацию магнитного момента на границе:

$$\hat{m} = \begin{pmatrix} m_z & m_x - im_y & 0 & 0 \\ m_x + im_y & -m_z & 0 & 0 \\ 0 & 0 & m_z & m_x + im_y \\ 0 & 0 & m_x - im_y & -m_z \end{pmatrix}$$

Линейное приближение

Вблизи температуры сверхпроводящего перехода можно воспользоваться линейным приближением, в

котором $g_{\uparrow\uparrow} = g_{\uparrow\downarrow} = 1$. Тогда для аномальных функций Грина получим четыре уравнения

$$\begin{aligned} iDf_{\uparrow\downarrow}''(z) - 2f_{\uparrow\downarrow} - 2\Delta &= 0 \\ iDf_{\downarrow\uparrow}''(z) - 2f_{\downarrow\uparrow} + 2\Delta &= 0 \\ iDf_{\uparrow\uparrow}''(z) - 2f_{\uparrow\uparrow} &= 0 \\ iDf_{\downarrow\downarrow}''(z) - 2f_{\downarrow\downarrow} &= 0. \end{aligned}$$

В общем виде их решение имеет вид

$$f_j(z) = A_j e^{i k_j z} + B_j e^{-i k_j z} + \frac{\Delta}{\varepsilon} (\delta_{j,\uparrow\uparrow} - \delta_{j,\downarrow\downarrow}).$$

Индекс j принимает значения $\uparrow\uparrow, \downarrow\downarrow, \uparrow\downarrow, \downarrow\uparrow$. Известные коэффициенты A_j, B_j должны быть найдены из линеаризованных граничных условий.

В случае ориентации ферромагнитного изолятора вдоль оси z граничные условия на соответствующей границе есть

$$\begin{aligned} dG_s f_{\uparrow\downarrow}' &= \pm 2NG_Q (e^{i\varphi} - 1) f_{\uparrow\downarrow} \\ dG_s f_{\downarrow\uparrow}' &= \mp 2NG_Q (e^{-i\varphi} - 1) f_{\downarrow\uparrow} \\ f_{\uparrow\uparrow}' &= f_{\downarrow\downarrow}' = 0. \end{aligned}$$

Здесь d длина сверхпроводника, знак в правой части равенств следует выбрать различным для правой и левой границы. Примечательно, что при $\varphi = 0$, то есть при отсутствии спинового смешивания, все правые части обращаются в нуль и граничные условия совпадают с таковыми для изолятора.

Развернём теперь ферромагнетик вдоль оси x . Тогда в сверхпроводнике возникнут триплетные корреляции, то есть граничные условия для триплетных компонент $f_{\uparrow\uparrow}$ и $f_{\downarrow\downarrow}$ аномальной функции Грина перестанут быть нулевыми:

$$\begin{aligned} dG_s f_{\uparrow\uparrow}' &= \pm NG_Q [(1 - \cos \varphi)(f_{\uparrow\uparrow} - f_{\downarrow\downarrow}) \\ &\quad + i \sin \varphi (f_{\uparrow\downarrow} - f_{\downarrow\uparrow})] \\ dG_s f_{\downarrow\downarrow}' &= \pm NG_Q [(1 - \cos \varphi)(f_{\downarrow\downarrow} - f_{\uparrow\uparrow}) \\ &\quad + i \sin \varphi (f_{\uparrow\downarrow} - f_{\downarrow\uparrow})] \\ dG_s f_{\uparrow\downarrow}' &= \pm NG_Q [(1 - \cos \varphi)(f_{\uparrow\downarrow} - f_{\downarrow\uparrow}) \\ &\quad + i \sin \varphi (f_{\uparrow\uparrow} - f_{\downarrow\downarrow})] \\ dG_s f_{\downarrow\uparrow}' &= \pm NG_Q [(1 - \cos \varphi)(f_{\downarrow\uparrow} - f_{\uparrow\downarrow}) \\ &\quad + i \sin \varphi (f_{\uparrow\uparrow} - f_{\downarrow\downarrow})]. \end{aligned}$$

Результаты

Решением уравнения Узалея вкпе с приведёнными выше граничными условия, были получены координатные зависимости аномальных функций Грина в сверхпроводнике

$$f_s = \frac{f_{\uparrow\downarrow} - f_{\downarrow\uparrow}}{2} \text{ и } f_t = \frac{f_{\uparrow\uparrow} + f_{\downarrow\downarrow}}{2}.$$

Рассмотрены случае параллельной ориентации ферромагнетиков вдоль оси z и перпендикулярной, в которой правый ферромагнетик ориентирован вдоль оси z , а левый – вдоль оси x .

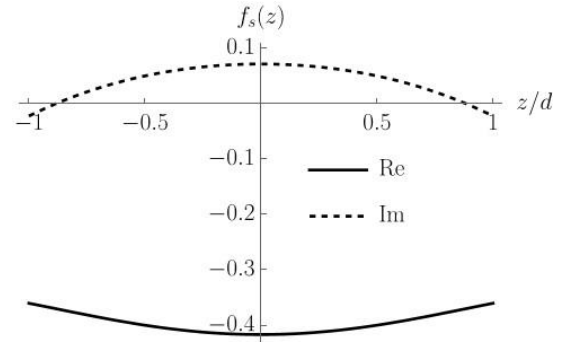


Рис.2. Синглетная аномальная функция Грина в сверхпроводнике при параллельной ориентации ферромагнетиков

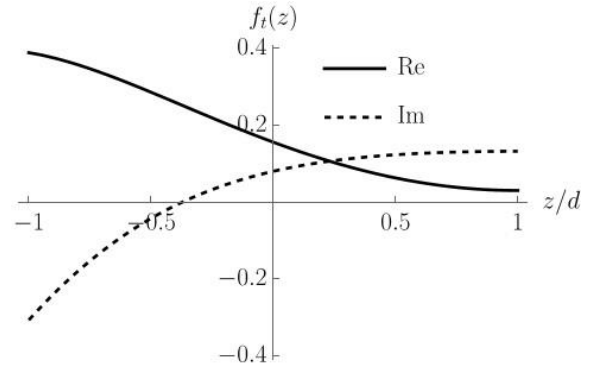


Рис.3. Триплетная аномальная функция Грина в сверхпроводнике при перпендикулярной ориентации ферромагнетиков

Заключение

В данной работе были решены уравнения Узалея в линейном приближении для гетероструктуры ферромагнитный изолятор-сверхпроводник-ферромагнитный изолятор. Вычисления проведены для параллельной и перпендикулярной ориентаций ферромагнитных изоляторов. В результате построены пространственные зависимости аномальных синглетных и триплетных функций Грина в сверхпроводнике. Показано, что при перпендикулярной ориентации ферромагнетиков в сверхпроводнике возникают триплетные корреляции, обычно отсутствующие в сверхпроводниках s -типа.

Список литературы

1. A. I. Buzdin, Proximity Effects in Superconductor-Ferromagnet Heterostructures, Rev. Mod. Phys. 77, 935 (2005).
2. S. Khlebnikov, Metastability and Bifurcation in Superconducting Nanorings, Phys. Rev. B 95, 174507 (2017).
3. P. Virtanen, F. S. Bergeret, and I. V. Tokatly, Magnetoelectric Effects in Superconductors Due to Spin-Orbit Scattering: Nonlinear σ -Model Description, Phys. Rev. B 104, 064515 (2021).
4. J. A. Ouassou, A. Pal, M. Blamire, M. Eschrig, and J. Linder, Triplet Cooper Pairs Induced in Diffusive S-Wave Superconductors Interfaced with Strongly Spin-Polarized Magnetic Insulators or Half-Metallic Ferromagnets, Sci Rep 7, 1932 (2017).

ЭЛЕКТРИЧЕСКАЯ ПРОВОДИМОСТЬ СЛОИСТЫХ СТРУКТУР ИНДИЙ – ПОЛИДЕФИНЕЛЕНФТАЛИД – ИНДИЙ ПРИ НИЗКИХ ТЕМПЕРАТУРАХ

Смазнова Х.Т.¹, Арутюнов К.Ю.¹, Завьялов В.В.¹,
Юсупов А.Р.², Карамов Д.Д.², Лачинов А.Н.²

¹Национальный исследовательский университет
«Высшая школа экономики»,
²Башкирский государственный
педагогический университет

Аннотация

Объектом нашего исследования являются гетероструктуры на базе полидифениленфталата (ПДФ). ПДФ – органический полимер, который в основном состоянии является диэлектриком, но в метастабильном состоянии, стимулированным внешним возмущением, обнаруживает электрическую проводимость. В настоящей работе в широком диапазоне температур экспериментально исследованы транспортные свойства тонкопленочных «сэндвичей» индий – ПДФ – индий. При температурах ниже ~3,4 К наблюдаются особенности, которые можно объяснить эффектом наведенной сверхпроводимости в тонкой пленке проводящего полимера, заключенной между двумя массивными сверхпроводниками (индий).

Введение

В настоящей работе рассматривается особый класс полимеров, который обладает таким эффектом, как стимулирование металлического состояния. В основном состоянии они являются широкозонными диэлектриками, но могут проявлять высокую электропроводность под влиянием таких внешних параметров, как механическое напряжение или электрическое поле. ПДФ относится к классу карбоциклических органических электроактивных полимеров с высокой химической и термической стабильностью, прозрачностью в видимом спектре и высокой механической прочностью [1,2]. ПДФ имеет аморфную структуру со степенью кристалличности не более 15%. Он характеризуется высокой химической стойкостью. При нормальных условиях ПДФ является широкозонным диэлектрическим материалом и характеризуется следующими параметрами: ширина запрещенной зоны ≈ 4.3 эВ, работа выхода электрона ≈ 4.2 эВ, сродство к электрону ≈ 2 эВ, потенциал первой ионизации ≈ 6.2 эВ. Однако в другом (метастабильном) состоянии, например, индуцированным внешним электрическим полем, система характеризуется ненулевой плотностью электронных состояний внутри запрещенной зоны. Глубина таких состояний увеличивается, если система принимает дополнительный электрон, что косвенно обеспечивает электропроводность вдоль цепи полимера [3,4]. В недавних работах [5-7] в тонкопленочных сэндвичах свинец – ПДФ – свинец при толщинах слоя полимера от нескольких десятков до нескольких сотен нанометров в широком температурном диапазоне наблюдалась типичная для металлов зависимость: падение

электрического сопротивления R с уменьшением температуры T . Но самым неожиданным результатом оказалось, что в области температур ниже ~ 8 К $R(T)$ и $V(I)$ характеристики четко проявляли особенности, типичные для сверхпроводников. Наблюдения были интерпретированы как наведение сверхпроводящих корреляций в тонкой электропроводящей пленке полимера за счет эффекта близости с массивным сверхпроводником (свинец). В тех структурах свинцовые электроды изготавливались методом термического напыления: т.е. верхний слой свинца формировался из раскаленных частиц металла, осаждаемых на полимер. Хотя ПДФ и обладает высокой химической и термической стабильностью, нельзя полностью исключить возможность «прожигания» тонкой пленки полимера, приводящего к формированию металлических закороток. Анализ срезов образцов методом просвечивающей электронной микроскопии [5-7] не дал однозначных результатов, которые позволили бы исключить такой тривиальный сценарий. Целью настоящей работы является проведение исследований электронного транспорта в широком температурном диапазоне в структурах сверхпроводник – ПДФ – сверхпроводник, где в качестве сверхпроводящего материала выбран индий: исключительно мягкий металл (при комнатных температурах). Соответственно, в такой системе с большой вероятностью можно исключить возникновение механического повреждения пленки полимера в процессе изготовления.

Методика

В качестве сверхпроводящего материала был выбран индий, являющийся типичным представителем сверхпроводников 1-го рода с критической температурой $T_c(\text{In}) \approx 3,4$ К. При комнатных температурах индий – исключительно мягкий металл. Как нижние, так и верхние электроды гетероструктур In-PDP-In изготавливались холодной прокаткой индия. Субмикронные пленки ПДФ получали центрифугированием полимера из раствора в циклогексаноне на твердой подложке из химического стекла, на поверхности которого предварительно был сформирован нижний индиевый электрод. Измерения температурных зависимостей сопротивления $R(T, I_{ac} = \text{const})$ проводились в 4-х контактной конфигурации на переменном токе с использованием фазочувствительного детектирования. Величина измерительного тока I_{ac} выбиралась таким образом, что её увеличение в ~ 10 раз не приводило к сколько-нибудь заметному смещению точки сверхпроводящего фазового перехода. Измерительная схема (рис.1) в обязательном порядке включала в себя развязку криогенной части цепи от цифровой электроники через малошумящие аналоговые предусилители тока и напряжения. Критическая температура T_c определялась как точка, когда эффективное сопротивление $R(T)$ упало в 2 раза по сравнению с нормальным R_N . Регистрация вольт-амперных характеристик (ВАХ) при фиксированной температуре $V(I_{dc}, T = \text{const})$ проводились на постоянном токе I_{dc} .

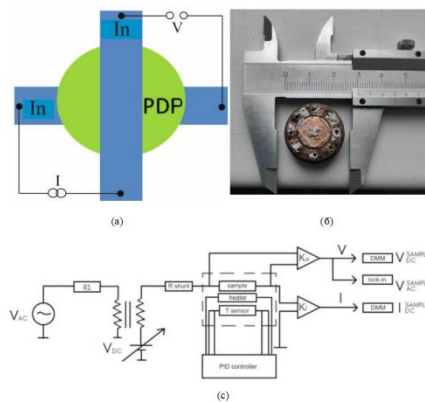


Рис.1. (а) Схематическое представление структуры в конфигурации «сэндвич». (б) Фотография держателя образца с образцом. (в) Схематика измерений

Результаты

Типичные зависимости сопротивления от температуры $R(T, I_{ac}=\text{const})$ для структуры In-PDP-In в конфигурации «сэндвич» приведены на рис.2. В широком температурном диапазоне система проявляет металлический ход сопротивления: уменьшение с понижением температуры (рис.2а) и резкое падение до инструментального нуля (рис.2б) ниже температуры сверхпроводящего перехода тонкопленочных индиевых электродов $T_c(\text{In}) \approx 3,4$ К.

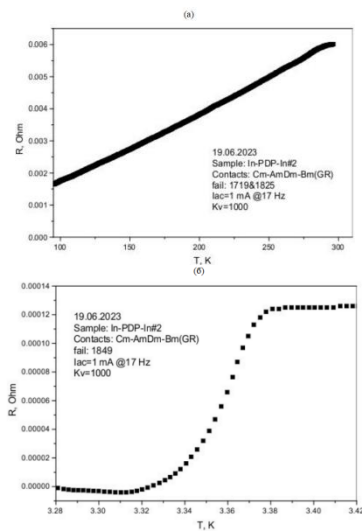


Рис.2. Зависимости сопротивления от температуры $R(T, I_{ac}=\text{const})$ для тонкопленочной слоистой гетероструктуры In-PDP-In#2 (а) в высокотемпературном диапазоне и (б) в области сверхпроводящего перехода

При фиксированной температуре ниже сверхпроводящего перехода ВАХ демонстрируют типичную зависимость для сверхпроводника: нулевое падение напряжения до некоего критического значения тока $I_c(T)$ (рис.3).

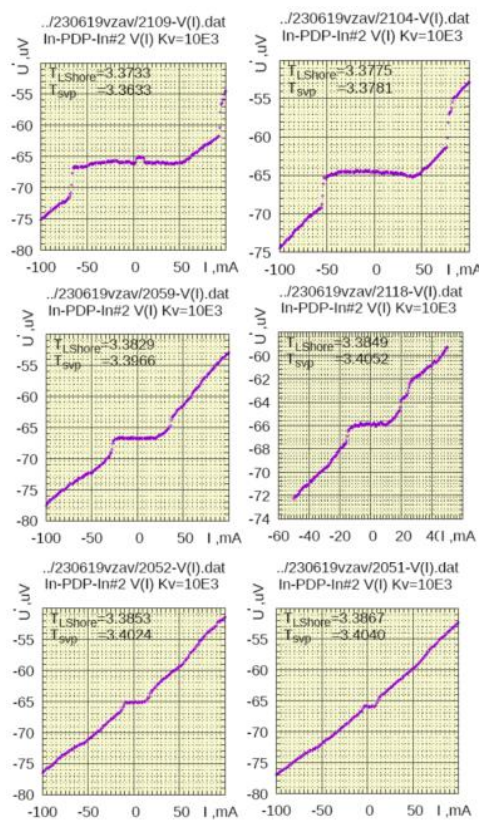


Рис.3. Вольт-амперные характеристики $V(I_{dc}, T=\text{const})$ для тонкопленочной слоистой структуры In-PDP-In#2 при фиксированных температурах ниже температуры сверхпроводящего перехода

Следует обратить внимание, что на экспериментальных ВАХ (рис.3) наблюдается некий гистерезис: критическое значение тока при его увеличении по абсолютной величине I_c^{UP} отличается от аналогичной величины при уменьшении I_c^{DOWN} . Это – обычное явление, связанное с локальным перегревом. Однако, ни те, ни другие зависимости критического тока от температуры не ложатся на зависимость $\sim (T_c - T)^{3/2}$, следующую из модели Гинзбурга – Ландау для квазиодномерного сверхпроводящего канала. Из анализа зависимостей на рис.4 можно сделать два вывода. Первый: канал(ы) протекания сверхтока через ПДФ не являются квазиодномерными [8-10]. Второй: сама форма зависимостей $I_c(T)$ (рис.4а) отличается от типичной для классических сверхпроводников.

$R(T)$ и $V(I)$ зависимости позволяют сделать предположение, что в слоистой гетероструктуре In-PDP-In присутствуют сверхпроводящие корреляции. Аналогичные результаты были уже получены ранее в системах Pb-PDP-Pb [5-7] и интерпретировались как проявление эффекта близости (proximity effect) за счет контакта со сверхпроводником (свинец). Качественно это наблюдение повторяется и в аналогичной системе на базе индия, что позволяет сделать осторожное предположение об универсальности явления. Как обсуждалось выше, в случае со свинцом открытым был вопрос о возможности формирования металлических закороток («проколов»), которые могли образоваться за счет проплавления

полимера при термическом напылении свинца. В исследованных в настоящей работе сэндвичах на базе индия термическое повреждение пленки ПДФ исключено, так как использовалась методика холодной прокатки индия. Также достаточно низка вероятность механического прокола в силу исключительной мягкости индия. В ближайшем будущем планируется проведение микроскопического анализа среза «сэндвичей» In-PDP-In методом растровой и просвечивающей электронной микроскопии.

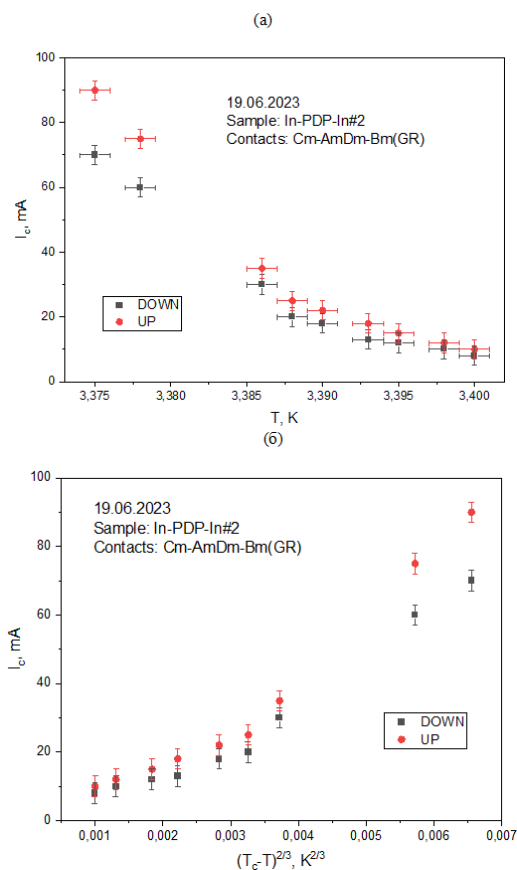


Рис.4. (а) Зависимость критического тока I_c сэндвича In-PDP-In#2 от температуры T . (б) Та же зависимость, построенная в координатах $(T_c - T)^{3/2}$

Заключение

Экспериментально исследовалась электрическая проводимость в тонкопленочных слоистых структурах In-PDP-In. При температурах ниже $\sim 3,4$ К во всех образцах на $R(T)$ и $V(I)$ зависимостях наблюдаются особенности, которые можно объяснить эффектом наведенной сверхпроводимости в тонкой пленке проводящего полимера, заключенной между двумя массивными сверхпроводниками (индием).

Работа поддерживалась программой сотрудничества «Зеркальные лаборатории» Национального Исследовательского Университета Высшая Школа Экономики и Башкирского Государственного Педагогического Университета им. М.В. Акмуллы.

Список литературы

1. A. N. Lachinov, V. M. Kornilov, T. G. Zagurenko, and A. Yu. Zherebov, On the high conductivity of non-conjugated polymers, *Journal of Experimental and Theoretical Physics*, 102(4), 640 (2006).
2. A. N. Lachinov, N. V. Vorob'eva, *Electronics of thin wideband polymer layers*, *Physics - Uspekhi* 49, 1223 (2006).
3. A.N. Lachinov, A.R. Tameev, A.R. Yusupov, A.V. Vannikov, Russ. J. Electrochem. Effect of interface in bilayer polydiphenyleneophthalide film on electron transport, 48(3), 316 (2012).
4. V. M. Kornilov, A. N. Lachinov, Electrical conductivity in the metal-polymer-metal system: The role of boundary conditions, *Journal of Experimental and Theoretical Physics*, 84, 833 (1997).
5. K. Yu. Arutyunov, K. A. Belyaev, V. V. Artemov, A. L. Vasil'ev, A. R. Yusupov, D. D. Karamov, and A. N. Lachinov. Transport Properties of Layered Heterostructures on the Base of a Conducting Polymer. *Physics of the Solid State* 64(12), 603 (2023).
6. K. Yu. Arutyunov, V.V. Artemov, A. L. Vasiliev, A. R. Yusupov, D. D. Karamov, A. N. Lachinov, Induced electric conductivity in organic polymers, *Beilstein J. Nanotechnology* 13, 1551 (2022).
7. K. Yu. Arutyunov, A. S. Gurski, S. D. Monakhova, P. V. Panarina, E. Ph. Pozdnyakova, D. E. Tsoi, A. R. Yusupov, D. D. Karamov, A. N. Lachinov, Low temperature electron transport in hybrid thin film nanostructures based on electroconducting polymer. *Proceedings of the RAS Ufa Scientific Centre* 1, 42 (2023).
8. R. Tidecks, *Current Induced Nonequilibrium Phenomena in Quasi-One-Dimensional Superconductors* (Springer, New York, 1990).
9. K. Y. Arutyunov, N. P. Danilova, A. A. Nikolaeva, Nonequilibrium galvanomagnetic properties of quasi-one-dimensional superconductors, *Phys. C: Supercond.* 235-240, 1967 (1994).
10. K. Y. Arutyunov, N. P. Danilova, A. A. Nikolaeva, Galvanomagnetic properties of quasi-one-dimensional superconductors, *J. Appl. Phys.* 76 (10), 7139 (1994).

ДИНАМИЧЕСКИЙ ЭФФЕКТ БЛИЗОСТИ В ГИБРИДНОЙ СТРУКТУРЕ СВЕРХПРОВОДНИК-ФЕРРОМАГНИТНЫЙ ДИЭЛЕКТРИК

Туркин Я.В.¹, Пугач Н.Г.^{1,2}

Национальный исследовательский университет «Высшая школа экономики»,
¹научно-учебная лаборатория квантовой наноэлектроники,
²департамент электронной инженерии
 МИЭМ НИУ ВШЭ

Аннотация

В работе рассмотрено взаимное влияние динамики намагниченности в ферромагнитном диэлектрике и динамики сверхпроводящих корреляций в слое прилегающего «грязного» сверхпроводящего металла.

Введение

Получение чистых бездиссипативных спиновых токов в сверхпроводниках открывает широкий потенциал для построения элементной базы сверхпроводящей спинтроники. Такие спинтронные элементы особенно актуальны, так как в них отсутствует джоулев нагрев, а для переключения логических состояний требуется значительно меньшие затраты энергии, чем в случае полупроводниковой технологии. Кроме того, близость сверхпроводника и ферромагнитного диэлектрика позволяет управлять спектром магнонов и обеспечивает ненулевой коэффициент связи между различными собственными спиновыми волнами. Эти свойства открывают возможности для построения сверхпроводящих магнонных устройств, таких как направленные магнонные ответвители [1] и магнонные логические элементы [2].

В данной работе, взаимное влияние сверхпроводника и ферромагнитного изолятора, рассматривается в рамках связанных уравнений Узаделя и Ландау-Лифшица-Гильберта.

Модель

Исследуемая гибридная структура ферромагнитный диэлектрик / сверхпроводник схематически изображена на рис. 1. Она состоит из слоя ферромагнитного диэлектрика (железо-иттриевый гранат) и слоя грязного сверхпроводника (ниобий).

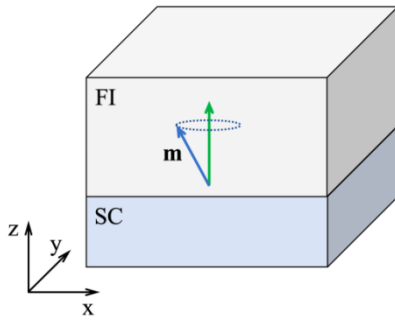


Рис.1. Схема исследуемой гибридной структуры ферромагнитный диэлектрик / сверхпроводник

Теоретическая модель динамики такой структуры, построена в формализме системы линеаризованных нестационарных уравнений Узаделя(1) и Ландау-Лифшица-Гильберта(2), которые описывают связанный линейный оклик сверхпроводника и ферромагнитного диэлектрика.

$$-\frac{i}{\hbar}\{\check{\rho}_4\partial_t, \check{g}^{(1)}\}_t + \hbar D\partial_z(\check{g}^{(1)} \circ \partial_z \check{g}^{(0)}) + \hbar D\partial_z(\check{g}^{(0)} \circ \partial_z \check{g}^{(1)}) = -i[\check{\Delta}^{(0)}, \check{g}^{(1)}], \quad (1)$$

где $\check{g}^{(1)}$ – линейная поправка к функции Грина, $\check{g}^{(0)}$ – функция Грина, соответствующая стационарной составляющей эффекта близости, $\check{\Delta}^{(0)}$ – стационарный параметр порядка.

$$\partial_t \mathbf{m}^{(1)} = -\gamma \left[\mathbf{M}^{(0)} \times \left(\mathbf{h}_{eff}^{(1)} + \frac{\alpha}{M_s} \partial_t \mathbf{m}^{(1)} \right) \right] - \gamma \left[\mathbf{m}^{(1)} \times \mathbf{H}_0^{(1)} \right], \quad (2)$$

где $\mathbf{m}^{(1)}$ – малая, нестационарная составляющая намагниченности, $\mathbf{M}^{(0)}$ – стационарная составляющая намагниченности, направленная по оси z, $\mathbf{H}_0^{(1)}$ – стационарная составляющая внешнего поля, $\mathbf{h}_{eff}^{(1)}$ – эффективное магнитное поле внутри диэлектрика, M_s – намагниченность насыщения, α – феноменологический параметр затухания. Учитывая периодический характер прецессии намагниченности, уравнение Узаделя может быть записано в представлении Фурье [3], что позволяет поучить стационарную задачу в частотном пространстве. Обратный эффект близости сверхпроводника с ферромагнитным диэлектриком может быть учтен при помощи нестационарного спинового матричного тока инжектируемого из магнитного слоя [4]. В целях упрощения модели, в граничных условиях для сверхпроводника учитывается только слагаемое первого порядка по углу спинового смешивания.

$$\check{g}^{(0)} \circ \partial_z \check{g}^{(1)} = iK[\check{m}_n, \check{g}^{(0)}], \quad (3)$$

где K – феноменологический угол спинового смешивания, определяющий силу обменного взаимодействия между электронами проводимости и намагниченностью в ферромагнитном диэлектрике.

Спиновый ток и наведенная намагниченность

Уравнения (1) и (2) были решены численно с граничными условиями (3). Полученная зависимость для наведенной намагниченности изображена внутри сверхпроводника [5] на рис. 2.

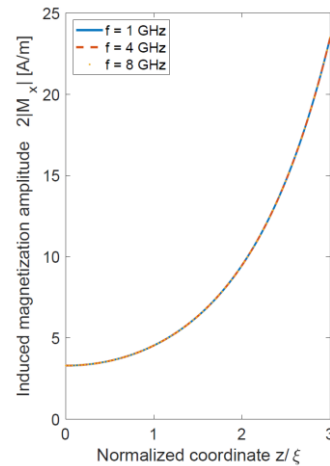


Рис.2. Распределение наведенной намагниченности в слое сверхпроводника

Можно видеть, что наведенная намагниченность не зависит от частоты, это обусловлено тем, что характер прецессии вектора намагниченности не изменяется с увеличением частоты, однако, при увеличении температуры и приближении частоты к удвоенной энергии сверхпроводящей щели следует ожидать увеличения роли эффекта спиновой накачки квазичастиц.

Спиновый ток в сверхпроводнике, возникающий при прецессии намагниченности в магнетике, изображен на рис. 3. Убывание спинового тока можно

объяснить аналогично спаду величины наведенной намагниченности: величина спинового тока пропорциональна плотности триплетных куперовских пар, а количество триплетных пар спадает с удалением от интерфейса из-за понижения эффективности динамической конверсии синглетных пар в триплетные.

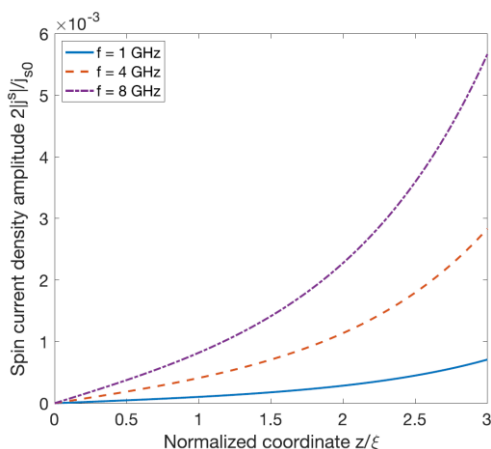


Рис.3. Распределение спинового тока внутри сверхпроводящего слоя при разных значениях частоты

Динамика квазичастиц в сверхпроводнике

Спиновая накачка квазичастиц также может вносить заметный вклад в динамический эффект близости при повышении температуры и/или приближении частоты прецессии магнитных моментов к энергии сверхпроводящей щели. На рис. 4 изображена зависимость величины наведенной намагниченности от частоты прецессии при разных значениях температуры T .

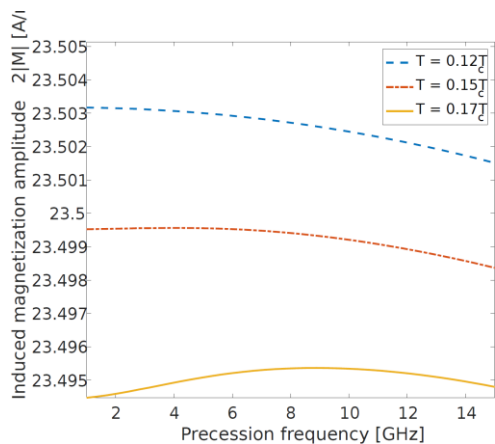


Рис.4. Наведенная намагниченность на интерфейсе между сверхпроводником и ферромагнитным диэлектриком

Появление максимума величины наведенной намагниченности при определенных значениях частоты и температуры можно объяснить конкуренцией между динамической конверсией синглетных пар в триплетные и спиновой накачкой квазичастиц. Данный эффект проявляется при повышении температуры, так как растет число квазичастиц. Эффект спиновой накачки квазичастиц может потенциально

проявляться в значительно большей степени при приближении частоты прецессии магнитных моментов к энергии щели.

Благодарности

Работа была выполнена при поддержке гранта РФФИ 23-72-00018 "Исследование неравновесных и граничных явлений в сверхпроводящих гибридных наноструктурах".

Заключение

Полученные результаты позволяют, оценить влияние взаимного динамического эффекта близости сверхпроводника и ферромагнитного диэлектрика на спиновый транспорт и магнетонный спектр.

Список литературы

1. Bobkova, I.V., et al. Magnon-cooperons in magnet-superconductor hybrids // *Comm. Mat.* - 2022. -№ 1. -Vol. 3.
2. Chumak A.V., Serga A.A., Hillebrands B. Magnon transistor for all-magnon data processing. // *Nature communications.* - 2014. -№ 1. -Vol. 5. -p. 4700.
3. Brinkman, A., et al. Microscopic nonequilibrium theory of double-barrier Josephson junctions // *Phys. Rev. B*, -2003. -Vol. 68. -№ 22. -p. 224513.
4. Eschrig, M., Cottet A., Belzig W., Linder J. General boundary conditions for quasiclassical theory of superconductivity in the diffusive limit: application to strongly spin-polarized systems // *New Journal of Physics.* - 2015. - Vol. 17. - p. 083037.
5. Turkin Y.V., Pugach N.G. Spin dynamics in superconductor/ferromagnetic insulator hybrid structures with precessing magnetization.// *Beilstein J. Nanotechnol.* -2023. -№ 14. -p. 233-239.

ИССЛЕДОВАНИЕ ВЛИЯНИЯ КВАНТОВО-РАЗМЕРНОГО ЭФФЕКТА НА СВЕРХПРОВОДЯЩИЙ ПЕРЕХОД В ТОНКИХ ПЛЕНКАХ АЛЮМИНИЯ

*Безымянных Д.Г., Пугач Н.Г.,
Седов Е.Г., Арутюнов К.Ю.*
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В данной работе выполнен расчёт изменения критической температуры в зависимости от толщины алюминиевой пленки, как для чистого случая без учета рассеяния электронов, так и с учетом рассеяния на немагнитных примесях.

Введение

С самого начала исследований в области сверхпроводимости было отмечено, что тонкие пленки могут проявлять критическую температуру перехода в сверхпроводящее состояние (T_c), которая значительно отличается от аналогичных значений для объ-

емных материалов [1]. Интересно отметить, что уменьшение толщины пленки может привести как к уменьшению (например, для ниобия), так и к увеличению (например, для алюминия) значения T_c . Несмотря на обширные экспериментальные данные и многообразие теоретических моделей, научное сообщество до сих пор не достигло единства мнений относительно природы этого явления. В данной работе мы представляем результаты теоретических и экспериментальных исследований высококачественных алюминиевых пленок на различных подложках.

Независимо от характеристик подложки и процесса формирования пленки мы выявляем одну и ту же закономерность: чем тоньше пленка алюминия, тем выше ее критическая температура. Мы предполагаем, что данное наблюдение связано с квантово-размерным эффектом [2].

Экспериментальные данные

Алюминиевые пленки были изготовлены с использованием либо электронно-лучевой эпитаксии (МВЕ), либо выращиванием на подложках из монокристалла арсенида галлия (GaAs) или сапфира. Толщина пленок варьировалась от 5 до 100 нм. Для всех пленок любой толщины параметр решетки алюминия равен 0,4053 нм, а направление роста равно (111).

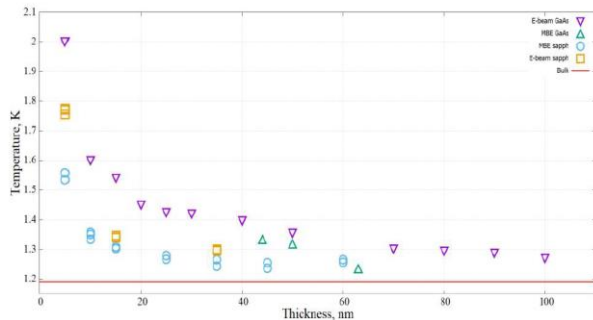


Рис.1. Критическая температура T_c в зависимости от толщины алюминиевых пленок, изготовленных различными методами и на разных подложках. Горизонтальная линия соответствует $T_c=1,19$ К для объемного алюминия [1]

Методика расчета

В данной работе рассчитывалась зависимость критической температуры сверхпроводящего перехода T_c от толщины тонкой пленки алюминия. В качестве метода для исследования квантово-размерного эффекта был выбран метод функций Грина, которые находятся из уравнений Горькова в рамках микроскопической модели сверхпроводимости БКШ. Для данной задачи о системе, ограниченной в пространстве, уравнения Горькова имеют следующий вид:

$$\begin{cases} (i\omega + \xi)G_\omega(r, r') + \Delta(r)F_\omega(r, r') = \delta(r - r') \\ (i\omega + \xi)F_\omega(r, r') + \Delta^*(r)G_\omega(r, r') = 0 \end{cases}$$

Здесь $\Delta(z)$ – величина энергетической щели, $G_\omega(r, r')$ и $F_\omega(r, r')$ – нормальная и аномальная

функция Грина соответственно в представлении Мацубары, ω – Мацубаровская частота.

Решение системы (1) необходимо подставить в уравнение самосогласования:

$$\Delta^*(r) = |g|T \sum_\omega F_\omega(r, r')$$

Так как $\Delta(r) \sim F_\omega(r, r')$, то можно сократить на Δ и получить уравнение вида:

$$f(T, D) = \frac{\hbar^2 \pi^2}{2m} \frac{\xi}{\lambda} = \text{const}$$

В данном уравнении ξ – длина когерентности, λ – константа электрон-фононного взаимодействия, \hbar – постоянная Планка, m – масса электрона, T – температура, D – толщина пленки. Величина $\frac{\hbar^2 \pi^2}{2m} \frac{\xi}{\lambda}$ не зависит от толщины и температуры образца, а является постоянной для данного материала.

Обсуждение

В качестве сверхпроводника был взят алюминий. Структура моделировалась при следующих параметрах: $T_c = 1,2$ К, $D = 5 - 100$ нм, $\xi = 100$ нм.

На рис.2. изображены рассчитанная (линии) и измеренная (символы) зависимость $T_c(D)$ для случая без рассеяния электронов. В качестве экспериментальных данных были взяты пленки, выращенные на подложках из сапфира (MBE sapph) и арсенида галлия (MBE GaAs).

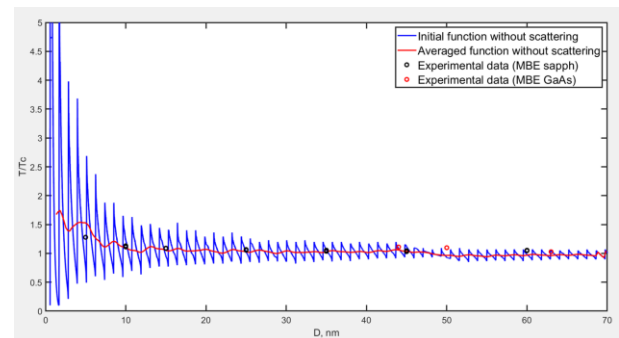


Рис.2. Зависимость критической температуры от толщины плёнки без учета рассеяния электронов на примесях

Из графиков следует что аналитический расчет был выполнен правильно так как с достаточно высокой точностью совпадает с экспериментальными данными.

Благодарности

Работа выполнена при поддержке проекта «Зеркальные лаборатории» Национального исследовательского университета «Высшая школа экономики» и Башкирского государственного педагогического университета им. М. Акмуллы.

Список литературы

1. A. I. Shal'nikov, Nature 9, 142 (1938).
2. K. Yu. Arutyunov, et. al., Phys. Status Solidi RRL 13, 1800317 (2019).

УПРАВЛЯЕМЫЙ ЭФФЕКТ БЛИЗОСТИ В СТРУКТУРЕ СВЕРХПРОВОДНИК- ГЕЛИКОИДАЛЬНЫЙ МАГНЕТИК

Лекомцев Н.В., Пугач Н.Г.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В работе изучается управляемый эффект близости под контролем спинового переключения в бислое сверхпроводник-геликоидальный магнетик, проводится сравнение расчётных значений критических параметров структуры с экспериментальными. Исследование проводится в рамках разработки модели сверхпроводникового спинового вентиля.

Введение

Повышенный интерес к устройствам сверхпроводниковой спинтроники вызван тем, что они являются перспективной альтернативой традиционным электронным компонентам – например, транзисторам – ввиду лучших характеристик. Сверхпроводниковые спиновые вентили позволят решить существующие проблемы современной микроэлектроники – сильный нагрев компонентов, высокое энергопотребление и технологическая сложность увеличения плотности их размещения на кристалле [1,2].

Целью работы является теоретическое описание и расчёт эффекта близости в структуре сверхпроводник-геликоидальный магнетик для его сравнения с результатами экспериментов. Исследование проводится в рамках разработки модели сверхпроводникового спин-вентиля.

В исследовании решаются следующие задачи: расчёт параметров эффекта близости в пленочной структуре с конкретными параметрами с целью сравнения его влияния на критические параметры для структур с двумя типами магнетика - парамагнетика и геликоидального магнетика.

В первом разделе описана модель исследуемой структуры, заданы параметры структуры. Во втором разделе описана методика проведения расчётов. В третьем разделе представлены результаты расчётов, их обоснование и сравнение с экспериментом.

Модель структуры

На рис.1 показано схематическое изображение исследуемой структуры – контакта сверхпроводящей пленки ниобия Nb с пленкой геликоидального магнетика силицида марганца MnSi. Оба слоя имеют конечную толщину d_s и d_f соответственно.

Силицид марганца - материал с неоднородной магнитной структурой, в нём реализуется геликоидальное магнитное упорядочение, что обуславливает появление дальнедействующих триплетных корреляций (LRTC), наличие которых позволяет куперовским парам в большем количестве покидать сверхпроводящий слой, чем при обычном эффекте близости. Использование магнетика с неоднородной магнитной

структурой позволяет выращивать контакты без буферных магнитных слоев с фиксированной намагниченностью, что значительно упрощает её изготовление. Сверхпроводящий слой изготовлен из ниобия, так как это материал имеет такие критические параметры, что сверхпроводящее состояние не разрушается при перемагничивании магнитного слоя [3 – 5].

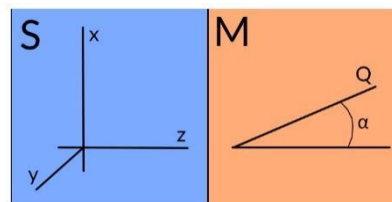


Рис.1. Структура исследуемого бислоя

Магнитный слой структуры характеризуется направлением спирального вектора Q , главным образом углом, который он образует с осью перпендикулярной плоскости контакта (осью OZ) – углом α , его модуль $Q = \frac{2\pi}{\lambda}$, где λ – период магнитной спирали (для MnSi $\lambda = 18$ нм). В MnSi вектор Q может быть направлен по кристаллографическим осям $[1\ 1\ 1]$ и $[\bar{1}\ \bar{1}\ \bar{1}]$ и эквивалентным им [6].

При помощи импульсов внешнего магнитного поля, противоположными по знаку и направленными параллельно плоскости контакта, можно изменять угол наклона вектора Q . Импульсы мощностью порядка 10^3 Э с длительностью порядка 10^2 пс вызовут переориентацию магнитной спирали, но при этом не разрушат сверхпроводящее состояние. Время переключения состояния сравнимо со временем переключения элементов традиционной спинтроники и составляет порядка нескольких нс [6].

Таблица 1. Параметры структуры

| Параметр | Значение |
|---|-------------------------|
| Критическая температура объёмного сверхпроводника T_c | 7,4 K |
| Длина когерентности для S-слоя ξ_s | 11 нм |
| Длина когерентности для F-слоя ξ_f | 4,2 нм |
| Толщина S-слоя d_s | 20 нм |
| Толщина F-слоя d_f | 20 нм |
| Величина обменного расщепления h | 11 мэВ |
| Величина вектора магнитной спирали Q | $0,349\ \text{нм}^{-1}$ |
| Параметр прозрачности γ_b | 0,7 |
| Параметр контакта γ | 0,7 |

Для геликоидальных магнетиков характерна доменная структура, как и для прочих магнитных материалов. Внутри каждого домена возникает своё направление вектора магнитной спирали Q из-за разных магнитных моментов, поэтому необходимо приложить внешнее магнитное поле (порядка 10 мэВ) для переориентации векторов магнитной спирали каждого домена вдоль поля. В таком случае угол

падения внешнего поля относительно поверхности пленки можно считать углом ориентации вектора \mathbf{Q} (относительно той же оси, что угол падения внешнего поля).

Основные параметры для расчёта структуры взяты из статьи Н.Г. Пугач [6] и работы П. Шрёдера [7]. Они представлены в таблице 1.

Методика исследования

В пленочных сверхпроводящих структурах реализуется диффузионный (грязный) предел, поэтому для расчётов удобно использовать транспортные уравнения Узалея (1), линеаризованные по аномальным частям функции Грина [8]

$$\begin{aligned} (D_{f,s}\nabla^2 - 2|\omega_n|)f_s &= -2\pi\Delta + 2i \cdot \text{sgn}(\omega_n)hf_t \\ (D_{f,s}\nabla^2 - 2|\omega_n|)f_t &= 2i \cdot \text{sgn}(\omega_n)hf_s, \end{aligned} \quad (1)$$

где f_s, f_t – синглетная и триплетная аномальные функции Грина соответственно, $D_{f,s}$ – коэффициенты диффузии соответствующего слоя, $\omega_n = \pi T(2n + 1)$ – мацубаровские частоты (максимальная частота ограничена частотой Дебая, которая для данной структуры соответствует температуре Дебая $T_D = 350 K$, откуда получаем, что $n = 0 \dots 12$), Δ – синглетный параметр порядка, $h(r) = h(\cos \cos(Qz), \sin \sin(Qz), 0)^T$ – вектор обменного поля с размерностью обменного расщепления.

Уравнения дополняются граничными условиями Куприянова-Лукичева (2) и (3) [9]. На границах структуры при $z = d_f$ и $z = -d_s$ и при наклоне вектора \mathbf{Q} на угол α они примут вид:

$$\frac{\partial}{\partial z} f_{s,t} = 0 \quad (2)$$

На границе контакта двух слоев при $z = 0$:

$$\begin{aligned} \xi_s \frac{\partial}{\partial z} f_{s,t}^S &= \gamma \xi_f \frac{\partial}{\partial z} f_{s,t} \\ f_{s,t}^S &= f_{s,t} - \gamma_b \xi_f \frac{\partial}{\partial z} f_{s,t} \end{aligned} \quad (3)$$

Для расчёта угловых зависимостей (от угла приложения θ внешнего магнитного поля, который отсчитывается и связывается с углом наклона вектора магнитной спирали α соотношением $\theta = \frac{\pi}{2} - \alpha$) критических параметров – поля и температуры – используются формулы (4) и (7) соответственно [10 – 12].

$$\frac{H_c(\theta)\sin\theta}{H_{c\perp}} + \left(\frac{H_c(\theta)\cos\theta}{H_{c\parallel}} \right)^2 = 1 \quad (4)$$

где $H_{c\perp}$ и $H_{c\parallel}$ – значения перпендикулярного и параллельного критического поля соответственно равные:

$$H_{c\perp} = \frac{\Phi_0}{2\pi 0,74^2 \xi_0^2} \left(1 - \frac{T}{T_c} \right) = h_{\perp} \left(1 - \frac{T}{T_c} \right) \quad (5)$$

$$H_{c\parallel} = \frac{\sqrt{3}\Phi_0}{\pi d_s 0,74 \xi_0} \left(1 - \frac{T}{T_c} \right)^{\frac{1}{2}} = h_{\parallel} \left(1 - \frac{T}{T_c} \right)^{\frac{1}{2}} \quad (6)$$

где ξ_0 – длина когерентности при нулевой температуре, Φ_0 – квант магнитного потока.

$$\frac{T_c(\theta)}{T_c(\theta=90^\circ)} = \frac{1 - H_{ext}\sin\theta h_{\perp}^{-1} - H_{ext}^2\theta h_{\parallel}^{-2}}{1 - H_{ext}\sin\theta h_{\perp}^{-1}} \quad (7)$$

где H_{ext} – внешнее магнитное поле.

Обсуждение результатов

П. Шрёдер в своей работе [7] экспериментально исследует пленочные структуры с двумя типами магнитных материалов: контакты сверхпроводника с парамагнетиком (Nb/CoSi) и с геликоидальным магнетиком (Nb/MnSi). Угловая характеристика критической температуры контакта с силицидом марганца MnSi пролегает ниже, чем для контакта с силицидом кобальта CoSi, и подавляется сильнее при малых значениях угла α .

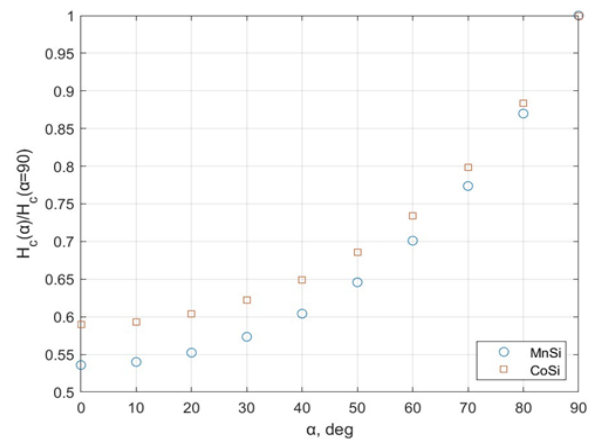


Рис.2. Угловая зависимость критического поля

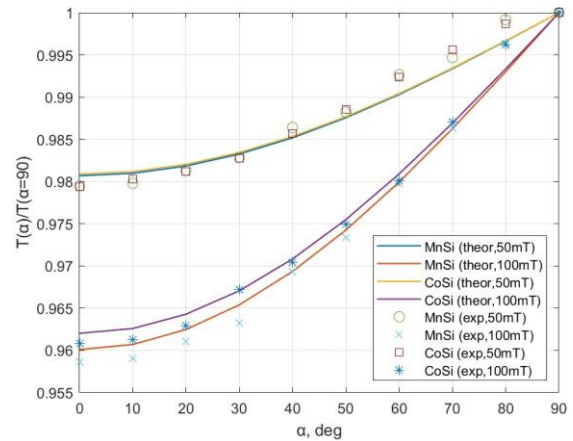


Рис.3. Угловые зависимости критической температуры: сплошными линиями – расчетные, точками – экспериментальные

На рис.2 и рис.3 представлены расчётные угловые характеристики критического поля и критической температуры соответственно. Наблюдается общая зависимость – с уменьшением угла уменьшаются значения критического поля и критической температуры, что согласуется с результатами экспериментов. На рис.3 видно, что при значении внешнего поля 50 мТл характеристики критической температуры сов-

падают для двух материалов, при значении поля, достаточном для переориентации векторов магнитных спиралей в доменах в MnSi вдоль поля (в данном случае 100 мТл), видно различие в характеристиках критической температуры – для MnSi наблюдается эффект близости, зависящий от угла – с уменьшением угла критические параметры подавляются сильнее, чем в CoSi, имеющем парамагнитную магнитную структуру.

Полученные расчётные угловые зависимости хорошо совпадают с экспериментальными данными.

Заключение

Использование геликоидального магнетика в качестве одного из функциональных слоев сверхпроводникового спин-вентили обосновано более выраженным эффектом близости, что подтверждает хорошее согласие расчетных и экспериментальных данных.

Благодарность

Авторы благодарят Научный фонд НИУ ВШЭ и программу сотрудничества «Зеркальные лаборатории» Национального Исследовательского Университета Высшая Школа Экономики и Башкирского Государственного Педагогического Университета им. М.В. Акмуллы.

Список литературы

1. Matthias Eschrig, Spin-polarized supercurrents for spintronic, Phys. Rev. Lett., 2016, 14c.
2. J. Linder, J.W.A. Robinson. Nature Phys. 11, 307 (2015).
3. Стишов С М, Петрова А Е "Геликоидальный зонный магнетик MnSi: магнитный фазовый переход" УФН 187 1365–1374 (2017)
4. Mühlbauer, S. et al. Skyrmion lattice in a chiral magnet. Science 323, 915–919 (2009).
5. Tonomura, A. et al. Real-space observation of skyrmion lattice in helimagnet MnSi thin samples. Nano Lett. 12, 1673–1677 (2012).
6. N.G. Pugach, M.O. Safonchik, V.I. Belotelov, T. Ziman, and T. Champel Phys. Rev. Applied 18, 054002
7. Schröder P. Implementierung einer winkelabhängigen Widerstandsmessung zur Untersuchung supraleitender Schichtsysteme // Wissenschaftliche Arbeit zur Erlangung des akademischen Grades Master of Science im Fach Physik // Technischen Universität Braunschweig im Institut für Physik der Kondensierten Materie. Braunschweig, 2023
8. T. Champel, M. Eschrig. Phys. Rev. B 71, 220506(R) (2005).
9. М.Ю. Куприянов, В.Ф. Лукичев. ЖЭТФ 94, 139 (1988).
10. Tinkham M. Effect of Fluxoid Quantization on Transitions of Superconducting Films, Phys. Rev. 129, 2413 (1963)
11. Yamafuji K., Kawashima T. and Irie F. On the angular dependence of the upper critical field in thin films. Physics Letters, 20(2), 1966
12. Harper F.E. and Tinkham M. The Mixed State in Superconducting Thin Films, Phys. Rev. 172, 441 (1968)

ОСОБЕННОСТИ ПОВЕДЕНИЯ ПЛОТНОСТИ СОСТОЯНИЙ НА ГРАНИЦАХ ГЕТЕРОСТРУКТУРЫ СВЕРХПРОВОДНИК ФЕРРОМАГНИТНЫЙ ИЗОЛЯТОР

Селезнёв Д.В., Пугач Н.Г.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе исследуются особенности плотности состояний (ДоС), вызванные влиянием ферромагнитного изолятора на сверхпроводниковую плёнку. Теоретическая модель основывается на численном решении уравнений Узаделя с граничными условиями, описывающие сильное спин зависимое отражение на границе интерфейса. Численно рассчитанные особенности ДоС на границах структуры могут быть экспериментально исследованы методами сканирующей туннельной микроскопией.

Введение

Развитие современных сверхпроводящих вычислительных устройств на квантовой логике нуждается в разработке элементов памяти на сверхпроводниках. Для достижения этой цели сегодня активно исследуются квантовые эффекты близости в гибридных SF структурах [1].

Одним из таких эффектов является обратный эффект близости. Впервые этот эффект наблюдался в работе [2]. Обратный эффект близости характеризуется возникновением индуцированной намагниченности в сверхпроводнике на расстоянии длины когерентности от ферромагнетика. Вклад в наведённую намагниченность вносят триплетные куперовские пары, которые образуются на границе раздела вследствие спин зависимого отражения, что приводит к нарушению симметрии пар [3]. В отличие от синглетных куперовских пар, полный спин триплетной пары равняется единице, поэтому могут возникать ненулевые проекции спина на ось намагниченности. В свою очередь это позволяет использовать этот эффект в устройствах сверхпроводниковой спинтроники, в которых управление сверхпроводящим спиновым током осуществляется через воздействие на спины электронов в триплетной паре.

Намагниченность в сверхпроводнике вблизи границы раздела S-FI также связано со спиновым расщеплением локальной плотности состояний (ДоС) [4]. Спектроскопические эксперименты в алюминиевых плёнках в контакте с халькогенидами европия, такими как EuO и EuS, продемонстрировали сверхпроводимость, с расщеплением спина в структуре плотности состояния. Теоретическая работа [5] объясняла этот эффект введением параметра, описывающего усредненное Зеемановское поле в сверхпроводнике, вызванное влиянием ферромагнитного изолятора. Аналогичный метод использовался для объяснения поведения ДоС в ряде экспериментов со структурами Eu/Al [6]. Тем не менее, такой теорети-

ческий подход не учитывает близость к границе раздела и подавление параметра порядка в сверхпроводнике.

С другой стороны, ДоС в сверхпроводниковом слое может быть сильно изменена спиновыми триплетными сверхпроводящими корреляциями, которые проявляются вблизи ферромагнетика. Данная работа основывается на самосогласованном расчёте обратного эффекта близости, который позволяет учесть недостатки предыдущих методов и вычислить особенности ДоС, как на границе интерфейса, так и в глубине сверхпроводника.

Модель

Для описания обратного эффекта близости в S/FI гетероструктуре, используются уравнения Узалея в параметризации Риккати ($\gamma, \tilde{\gamma}$).

$$D(\partial_z^2 \gamma + 2(\partial_z \gamma) \tilde{\gamma} N(\partial_z \gamma)) + i(c_{11} \gamma - \gamma d + \gamma c \gamma - c_{12}) = 0$$

$$D(\partial_z^2 \tilde{\gamma} + 2(\partial_z \tilde{\gamma}) \gamma \tilde{N}(\partial_z \tilde{\gamma}) - i(c_{22} \tilde{\gamma} - \tilde{\gamma} a + \tilde{\gamma} b \tilde{\gamma} - c_{21}))$$

Здесь D – коэффициент диффузии, а коэффициенты c_{11} , c_{12} , c_{21} , c_{22} определяются из следующей матрицы 2×2 :

$$(c_{11} \ c_{12} \ c_{21} \ c_{22}) = (\epsilon + i\eta) \hat{\epsilon}_3 + \hat{\Delta}(z)$$

$\hat{\Delta}$ – параметр порядка, ϵ – энергия квазичастицы, η – параметр неупругого рассеивания, $\hat{\epsilon}_3$ – третья матрица Паули.

К уравнения добавляются граничные условия, описывающие спиновое отражение на границе, которое характеризуется углом спинового смешивания φ [7].

Параметр порядка рассчитывается на основе следующего соотношения, которое справедливо в режиме слабой связи $n_0 \lambda \ll 1$, где n_0 – плотность состояний на уровне Ферми в нормальном состоянии, λ – константа спаривания из теории БКШ, а Δ_0 – параметр порядка для объемного сверхпроводника.

$$\frac{\Delta(z)}{\Delta_0} = \frac{1}{2} n_0 \lambda \int_0^{\sinh(1/n_0 \lambda)} d\left(\frac{\epsilon}{\Delta_0}\right) [(N\gamma)_{12} - (N\gamma)_{21} - (\tilde{N}\tilde{\gamma})_{12}^* + (\tilde{N}\tilde{\gamma})_{21}^*] \cdot \tanh\left(\frac{\pi}{2e^c} \frac{\epsilon}{T/T_c}\right)$$

Для расчёта плотности состояния используется следующее выражение:

$$n(\epsilon) = \frac{1}{2} \text{Tr} \left(\text{Re}(N(1 + \gamma \tilde{\gamma})) \right)$$

Где N и \tilde{N} – нормировочные матрицы, зависящие от параметров Риккати:

$$N = (1 - \gamma \tilde{\gamma})^{-1}, \quad \tilde{N} = (1 - \tilde{\gamma} \gamma)^{-1}$$

Обсуждение

Самосогласованное численное решение уравнений Узалея из предыдущей главы позволяет вычислить параметры Риккати S-FI структуры и плотность состояний. В качестве сверхпроводника в расчёте использовалась алюминиевая плёнка. Параметр неупругого рассеивания $\eta = 0.02$. Температура была выбрана равной нулю, а критической была $T_c = 1.2$ К для Al. Длина когерентности сверхпроводника $\xi \approx 100$ нм была оценена исходя из значений T_c и D_s in в грязном пределе на основе следующего выражения

$$\xi = \sqrt{\frac{\hbar D_s}{2\pi k_B T}}$$

Все расчёты приведены для указанных значений параметров. На рисунках 1-2 толщина сверхпроводникового слоя $L = 120$ нм.

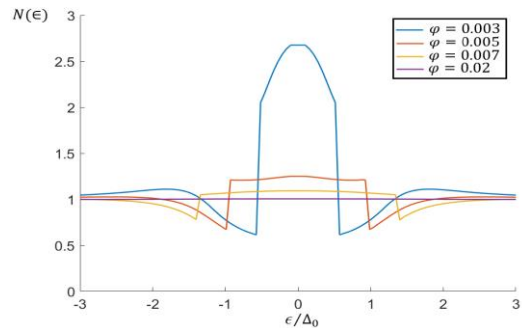


Рис.1. ДоС в слое S вблизи интерфейса S-FI для различных значений угла φ

На рис.1 показано, что вблизи границы раздела S-FI появляются только характерный нулевой пик (ЗБП) со слабыми признаками Зеемановского расщепления. При $\varphi > 0,02$ сверхпроводимость полностью подавляется. ЗБП почти полностью занимает область спектра, что является следствием триплетных сверхпроводящих корреляций, которые преобладают на границе раздела FI. При $\varphi = 0,005$ проявляется сдвиг пиков относительно энергетической щели, что является следствием расщепления спина. Эта структура ДоС может быть получена экспериментально в измерениях характеристик dI/dV тонкого туннельного барьера FI.

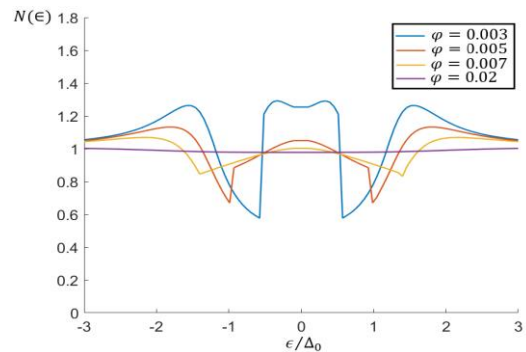


Рис.2. ДоС в слое S вблизи свободной границы раздела для различных значений угла φ

На рис.3 показано наличие ЗБП на внешней границе раздела S. Для больших φ также наблюдается полное подавление сверхпроводимости в тонком слое S толщиной порядка длины когерентности. Для $\varphi = 0,003$ проявляется особенность в виде удвоение нулевого пика, что может свидетельствовать о наличии Зеемановского расщепления в системе.

На свободной границе раздела наблюдается немонотонность ДоС на уровне Ферми в зависимости от толщины L. Эта немонотонная зависимость подробно представлена на рис.3.

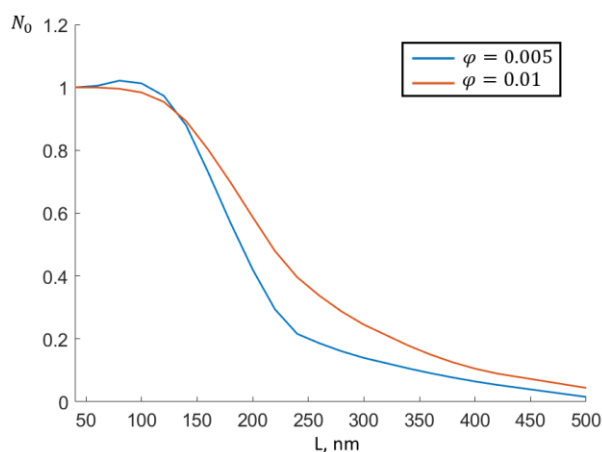


Рис.3. Зависимость ДоС на уровне Ферми от толщины сверхпроводящего слоя для $\varphi=0,005$ и $\varphi=0,01$ на свободном интерфейсе

Из рис.3 следует, что зависимость ДоС вблизи свободной границы раздела п от толщины слоя S может быть немонотонной (например, при $\varphi = 0,005$) или монотонной при большем φ . Сначала, при $L = 40$ нм, сверхпроводящие корреляции подавляются, затем появляется ЗБП примерно от 60 нм до 150 нм. Далее значение ДоС для нулевой энергии уменьшается до нуля и на длине около 500 нм соответствует случаю объемного сверхпроводника. Интересно, что максимальный пик приходится на длину около полутора длин когерентности. Обнаруженная зависимость ДоС показывает, что с увеличением угла спинового перемешивания сверхпроводимость в блисе подавляется сильнее, в то время как ЗБП становится менее выраженным. Мы наблюдали аналогичное немонотонное поведение для индуцированного намагничивания в нашей предыдущей работе [8].

Выводы

Таким образом, в этой работе исследованы особенности плотности состояния, вызванные расщеплением спина в S-слое. Мы показали, что даже при малых значениях параметра φ возникает значительное конвертация синглетных пар в триплетные. Согласно нашим выводам, индуцированное намагничивание и пик нулевого смещения (ЗБП) могут возникать даже в том случае, если в слое ферромагнитного изолятора (FI) отсутствует значительная спиновая поляризация или если связь с границей раздела FI не очень сильная.

Обнаруженное немонотонное поведение плотности состояния в зависимости от длины сверхпроводящего слоя демонстрирует, что структура сверхпроводящих корреляций чувствительна к толщине плёнки. Поэтому решения для толстых сверхпроводников могут сильно отличаться от тонких, и в них могут возникать совершенно другие особенности ДоС. Поскольку вместо приближения среднего обменного поля мы используем самосогласованный численный расчёт полученные результаты являются более точными.

Особенности плотности состояния на границе с вакуумом могут быть измерены экспериментально, используя методы сканирующей электронной микроскопии. Эти исследования могут быть важны для разработки новых сверхпроводящих спинтронных устройств.

Благодарности

Авторы благодарят за финансовую поддержку Научный фонд НИУ ВШЭ и Программу фундаментальных исследований НИУ ВШЭ (ТЗ-15).

Список литературы

1. Селезнёв Д.В., Пугач Н.Г., Львов Б.Г.//Наноиндустрия, 2021, том. 14(107), с. 572-573.
2. Tokuyasu T., Sauls J.A., Rainer D. Proximity effect of a ferromagnetic insulator in contact with a superconductor // Phys. Rev. B. 1988. Vol. 38, № 13. P. 8823–8833.
3. Bergeret F.S., Volkov A.F., Efetov K.B. Induced ferromagnetism due to superconductivity in superconductor-ferromagnet structures // Phys. Rev. B - Condens. Matter Mater. Phys. 2004. Vol. 69, № 17. P. 174504.
4. Yagovtzev V.O. et al. Magnetization in superconductor-ferromagnetic metal bilayers caused by the inverse proximity effect // Physics of Metals and Metallography. 2021. Vol. 122, № 9. P. 908–916.
5. Bobkova I.V., Bobkov A.M. Reconstruction of the DoS at the end of a S/F bilayer // Pis'ma JETP. 2019. Vol. 109, № 1. P. 57–62. [Pisma v ZhETF 2019. Vol. 109, № 1–2. P. 61–62].
6. 44. Bobkova I. V, Bobkov A.M., Belzig W. Signatures of spin-triplet Cooper pairing in the density of states of spin-textured superconductor-ferromagnet bilayers // New J. Phys. 2019. Vol. 21, № 4. P. 043001.
7. Селезнёв Д.В., Яговцев, В.О., Туркин Я. В., Пугач Н.Г., Екомасов Е.Г., Львов Б.Г.// Физика Металлов и Металловедение, 2023, том. 124(2), с. 196-203.
8. Yagovtzev V.O., Pugach N.G., Eschrig M. The inverse proximity effect in strong ferromagnet-superconductor structures // Supercond. Sci. Technol.

**СПИНОВОЕ РАССЕЯНИЕ
В ГЕТЕРОСТРУКТУРАХ
СВЕРХПРОВОДНИК-ФЕРРОМАГНЕТИК**

Милютин Д.П., Пугач Н.Г.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В ходе работы решены уравнения Узаледа для гетероструктуры сверхпроводник/ферромагнетик при различных видах спинового рассеяния. Получены численные решения и аналитические для линеаризованного вида. Проанализированы результаты и их зависимость от различных видов спинового рассеяния.

Введение

Одной из характерных величин для куперовских пар в сверхпроводнике является длина когерентности ξ – характерный масштаб, на котором происходит изменение параметра порядка ψ (волновой функции сверхпроводящего конденсата). Как следует из микроскопической теории сверхпроводимости, параметр ξ определяет размер электронной (куперовской) пары и определяется выражением $\xi_{S(M)} = \sqrt{\frac{D_{S(M)}}{2\pi T_c}}$, где $D_{S(M)}$ – постоянная диффузии электронов в сверхпроводнике (ферромагнетике), T_c – критическая температура. Особенно важную роль длина когерентности играет в случае хорошего контакта между нормальным металлом N и сверхпроводником S . Куперовские пары имеют возможность проникать в металл, что приводит к появлению сверхпроводящих корреляций в нем [1]. Такое поведение называется эффектом близости.

Одним из интереснейших примеров эффекта близости является гетероструктура сверхпроводник S и ферромагнетик F . Прежде всего это связано с противоположностью двух явлений – синглетной (электронные пары антисимметричны по спину) сверхпроводимости и ферромагнетизма, где электроны стараются выстроиться спинами параллельно обменному полю [2]. Основной особенностью эффекта близости в S/F структуре является поведение параметра порядка – он одновременно затухает и осциллирует. Такое необычное поведение влечет за собой немонотонную зависимость критической температуры и критического тока от ширины F слоя, а также так называемый π – контакт в $S/F/S$ структурах. Наряду с этим, также наблюдаются затухающие осцилляции функции плотности состояний от ширины F слоя.

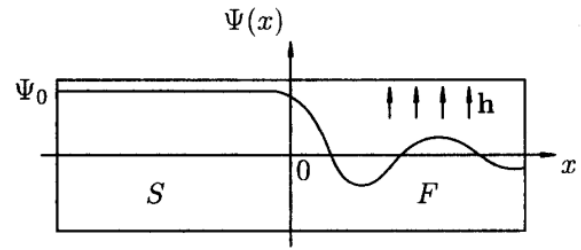


Рис.1. Схематическое изображение поведения параметра порядка вблизи SF-границы; h – обменное поле в ферромагнетике [1]

Причиной таких явлений является необычный механизм разрушения куперовских пар. Помимо обменного взаимодействия в F слое, которое стремится сориентировать спины электронов в одном направлении, влияет еще и спиновое рассеяние [3]. Всего есть три типа рассеяния – спин-орбитальное (τ_{so}), которое возникает из-за взаимодействия спина с магнитным полем, возникающем при движении электрона. Спин-флип, вдоль обменного поля, (τ_z) возникающее при перевороте спина рассеянного электрона, так как при этом меняется его энергия. Спин-флип, перпендикулярное обменному полю (τ_x) [4]. Спин-орбитальный механизм может разрушить осцилляции критической температуры и тока, в то время как спин-флип лишь изменяет их. Задачей данной работы является детальное исследование того, как это происходит.

Модель

Для описания эффекта близости в S/F гетероструктуре, при отсутствующем внешнем магнитном поле, используются уравнения Узаледа. Будем использовать следующую конструкцию.

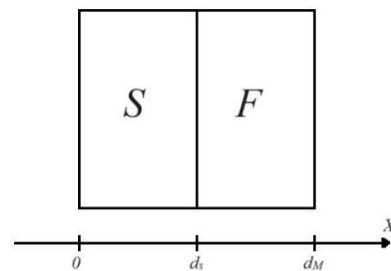


Рис.2. Схематичная модель гетероструктуры

В общем случае нелинейные, но вблизи критической температуры их можно линеаризовать с учетом малости сверхпроводящих корреляций. В них входят нормальная G и аномальная F функция Грина (пропагатор и коррелятор соответственно).

$$-\frac{D_S}{2} \nabla_r \cdot (G_{\uparrow\downarrow, \omega} \nabla_r F_{\uparrow\downarrow, \omega} - F_{\uparrow\downarrow, \omega} \nabla_r G_{\uparrow\downarrow, \omega}) + \omega F_{\uparrow\downarrow, \omega} = \Delta G_{\uparrow\downarrow, \omega}$$

Для S слоя,

$$-\frac{D_M}{2} \nabla_r \cdot (G_{\uparrow\uparrow, \omega} \nabla_r F_{\uparrow\downarrow, \omega} - F_{\uparrow\downarrow, \omega} \nabla_r G_{\uparrow\uparrow, \omega}) + (\omega + iE_{ex} + \tau_z^{-1} G_{\uparrow\uparrow, \omega} + (\tau_x^{-1} + \tau_{so}^{-1}) G_{\downarrow\downarrow, \omega}) F_{\uparrow\downarrow, \omega} + (\tau_x^{-1} - \tau_{so}^{-1}) G_{\uparrow\uparrow, \omega} F_{\downarrow\downarrow, \omega} = 0$$

для F слоя,

где $\omega = \omega_n = \pi T(2n + 1)$ – мацубаровские частоты с $n = 0, \pm 1 \dots$; T – температура; $D_s(D_M)$ – постоянная диффузии электрона для сверхпроводящего (несверхпроводящего) материала; $I_s(I_M)$ – область, соответствующая сверхпроводящим (ферромагнитным) слоям; $\tau_x, \tau_z, \tau_{so}$ – характерные времена спин-флип взаимодействий и спин-орбитального рассеяния, соответственно; E_{ex} – энергия обменного взаимодействия; Δ – параметр сверхпроводящего порядка [2].

Функции Грина связаны следующим условием нормировки $G^2 + FF^* = 1$.

Линейное приближение

Уравнения могут линейризованы вблизи критической температуры T_c либо при непрозрачной границе перехода. В первом случае это объясняется следующим – сверхпроводимость слабая, параметр порядка мал, а значит корреляции малы, то есть можем принять F малым. Во втором, опять же сверхпроводимость слабо проникает в ферромагнетик, а значит сверхпроводящие корреляции в нем малы.

Направим ось X перпендикулярно плоскости контакта. Тогда получим следующие уравнения

$$\left(|\omega| - \frac{D_s}{2} \frac{\partial^2}{\partial x^2} \right) F_{\uparrow\downarrow(\uparrow\downarrow), \omega}(x) = \Delta(x)$$

для S слоя, и

$$\left(|\omega| - \frac{D_M}{2} \frac{\partial^2}{\partial x^2} \pm iE_{ex} \operatorname{sgn}(\omega) + \tau_z^{-1} + 2\tau_x^{-1} \right) F_{\uparrow\downarrow(\uparrow\downarrow), \omega}(x) + (\tau_{so}^{-1} - \tau_x^{-1}) \left(F_{\uparrow\downarrow(\uparrow\downarrow), \omega}(x) - F_{\downarrow\uparrow(\uparrow\downarrow), \omega}(x) \right) = 0,$$

для F слоя.

Данные уравнения дополняются граничными условиями Куприянова-Лукичева [5].

$$\gamma \xi_s \left(\frac{\partial F_s}{\partial x} \right)_{d_s} = \xi_M \left(\frac{\partial F_M}{\partial x} \right)_{d_s},$$

$$F_M(d_s) = F_M(d_s) - \xi_M \gamma_B \left(\frac{\partial F_M}{\partial x} \right)_{d_s},$$

где $\gamma = \frac{\rho_M \xi_M}{\rho_s \xi_s}$ – параметр, определяющий величину градиента функции Грина в ферромагнетике вблизи границы: чем больше проводимость в сверхпроводнике по сравнению с ферромагнетиком, тем больше градиент G ; $\rho_{s(M)}$ – удельное сопротивление в сверхпроводнике (ферромагнетике); $\gamma_B = \frac{R_B}{\rho_s \xi_1}$ – безразмерный параметр, описывающий границу; R_B – сопротивление границы [6].

Обсуждение

При проникновении в ферромагнетик, сверхпроводящие корреляции осциллируют и затухают с характерными длинами ξ_{F_1} и ξ_{F_2} соответственно. Общую длину когерентности в ферромагнетике можно записать как $\frac{1}{\xi_F} = \frac{1}{\xi_{F_1}} + i \frac{1}{\xi_{F_2}}$ [7]. В случае большой обменной энергии и незначительных магнитных рассеяний получаем $\xi_{F_1} = \xi_{F_2} = \sqrt{\frac{\hbar D_M}{E_{ex}}}$. Но, когда время спин-флип рассеяния становится довольно маленьким, затухание ξ_{F_1} становится короче затухания осцилляций $\xi_{(F_2)}$ и сверхпроводимость разрушается быстрее.

Благодарности

Работа выполнена в рамках студенческого проекта №1607 “Спиновое рассеяние в гетероструктурах сверхпроводник-ферромагнетик” МИЭМ НИУ ВШЭ. Авторы благодарят за финансовую поддержку Научный фонд НИУ ВШЭ и Программу фундаментальных исследований НИУ ВШЭ (ТЗ-15).

Список литературы

1. Шмидт В. В. Введение в физику сверхпроводников. – 1982.
2. Кушнир В. Н. Сверхпроводимость слоистых структур. – 2010.
3. Faure M. et al. Striking properties of Superconductor/Ferromagnet structures with spin-dependent scattering //arXiv preprint cond-mat/0601707. – 2006.
4. Gusakova D. Y. et al. Density of states in SF bilayers with arbitrary strength of magnetic scattering //Journal of Experimental and Theoretical Physics Letters. – 2006. – Т. 83. – С. 327-331.
5. Golubov A. A., Kupriyanov M. Y. The current phase relation in Josephson tunnel junctions //Journal of Experimental and Theoretical Physics Letters. – 2005. – Т. 81. – С. 335-341.
6. Karabassov T. et al. Density of states in the presence of spin-dependent scattering in SF bilayers: a numerical and analytical approach //Beilstein journal of nanotechnology. – 2022. – Т. 13. – №. 1. – С. 1418-1431.
7. Obobnov V. A. et al. Thickness dependence of the Josephson ground states of superconductor-ferromagnet-superconductor junctions //Physical review letters. – 2006. – Т. 96. – №. 19. – С. 197003.

МЕТОДИКА ИЗГОТОВЛЕНИЯ И ФИЗИЧЕСКИЕ СВОЙСТВА ТУННЕЛЬНЫХ ПЕРЕХОДОВ МИКРОННЫХ РАЗМЕРОВ

*Позднякова Е.Ф.^{1,4}, Арутюнов К.Ю.^{1,3},
Тарасов М.А.⁴, Чекушкин А.М.⁴, Маркина М.А.^{2,4}*
Национальный исследовательский университет
«Высшая школа экономики»,

¹*департамент электронной инженерии*
МИЭМ НИУ ВШЭ,

²*базовая кафедра «Нанoeлектроника и фотоника»*
при Институте радиотехники и электроники им.
В.А. Котельникова РАН,

³*Институт физических проблем*
им. П.Л. Капицы РАН,

⁴*Институт радиотехники и электроники*
им. В.А. Котельникова РАН

Аннотация

В работе обсуждаются методики изготовления, первичного микроскопического анализа и тестирования микроструктур на базе сверхпроводников, изоляторов и нормальных металлов. Туннельные гибридные системы $\text{Cu-AIO}_x\text{-Al}$ предназначены для исследования неравновесных и граничных явлений в сверхпроводниках.

Введение

Реальные нанoeлектронные устройства представляют из себя гибридные структуры, включающие различные материалы. В таких системах обязательно присутствуют границы, что ведёт к возникновению разнообразных неравновесных и граничных явлений. С учётом тенденции к миниатюризации, большая часть такой системы может превратиться в "границу", поведение которой требует учёта указанных эффектов. Центральная научная задача данного исследования состоит в выявлении основных закономерностей неравновесного электронного транспорта в различных гибридных нанoeлектронных системах на основе сверхпроводников. Экспериментальное и теоретическое исследование структур мезоскопических размеров, включающих сверхпроводники (С), изоляторы (И) и несверхпроводники (Н), является главной задачей исследования в рамках проекта, включающее изучение интерферометров Ааронова-Бома [1] и эффектов релаксации квазичастичных возбуждений [2-3].

В работе описывается принцип формирования тестовых $\text{Al-AIO}_x\text{-Cu}$ гетероструктур с использованием таких методов, как вакуумное напыление и оптическая литография, первичный микроскопический анализ методами растровой электронной и оптической микроскопии.

Описание конструкции чипа для исследования тестовых гетероструктур

Для возможности проведения измерений, необходимых для изучения электрофизических свойств гибридных наноструктур, с использованием криогенных зондов при сверхнизких температурах (жидкого гелия и ниже) по четырехзондовой схеме, поддержи-

ваемой большей частью измерительного оборудования, была выбрана топология, которая используется для изготовления тестовых образцов в Институте радиотехники и электроники Российской академии наук (ИРЭ РАН). Топология рассчитана на чип размером 15×24 мм, на которой представлены 14 туннельных НИС/СИС переходов разных размеров от 1,5 до 5 мкм в диаметре (отмечены зелёной стрелкой на рис.1) с 16 контактными площадками (рис.1).

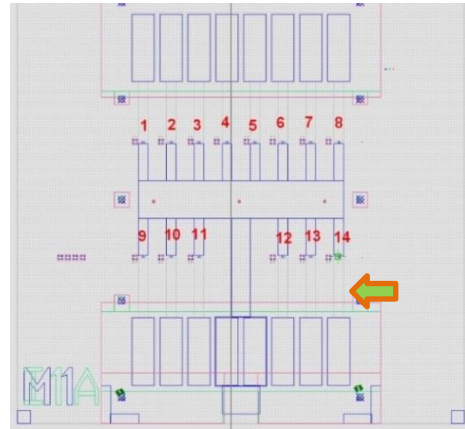


Рис.1. Схема чипа 15×24 мм с 14 тестовыми структурами

Топологию данных тестовых гетероструктур $\text{Cu-AIO}_x\text{-Al}$ можно будет использовать при дальнейшем дизайне магнитного интерферометра Ааронова-Бома [4], т.к. предполагается, что параметры туннельных контактов можно будет свободно масштабировать в сторону их уменьшения до суб-микрометровых размеров с помощью электронной литографии.

Изготовление НИС переходов: условия и технологический маршрут

Для изготовления тестовых гетероструктур $\text{Cu-AIO}_x\text{-Al}$ в качестве подложек использовались пластины стандартного размера 24×24 мм из монокристаллического кремния без дополнительного оксидного слоя на поверхности, который при низких температурах является диэлектриком. Перед работой подложки проходят очистку в органическом растворителе (ацетоне), промываются дистиллированной водой и высушиваются в струе сухого воздуха. Т.к. для формирования СИН переходов используется технология плазмохимического травления во избежание травления материала непосредственно подложки после чистки на них наносится защитный слой Al_2O_3 толщиной 100 нм методом магнетронного напыления.

На рис.2 показан порядок технологических действий при изготовлении гибридных структур. Была предложена новая технологическая карта, подразумевающая наличие разрыва вакуума между процедурами напыления слоёв меди и оксида алюминия, при этом считается, что из-за наличия «колодцев» в толстом слое изоляции, пленка оксида алюминия должна рваться, вследствие чего НИС переходы будут только на дне «колодцев».

Для реализации геометрии базового медного (Н) электрода под дальнейшую «взрывную» литографию на подложке формируется маска из фоторезиста ФП3515 толщиной ~ 1.5 мкм с использованием установки фотолитографии Karl Suss MA-150. Далее для устранения возможных органических загрязнений проводится чистка поверхности образца в кислородной плазме в течение 35 с, 50 Вт в установке плазмохимического травления March Instruments Inc. Jupiter II. Затем производится магнетронное напыление слоя меди в камере напыления Leybold AG Z-400 при остаточном давлении газов в камере $5 \cdot 10^{-6}$ мбар. После напыления производится удаление пленки фоторезиста с лишним слоем металла на ней в диметилформамиде в течение 1 часа при температуре 60 °С, а затем в ультразвуковой ванне в течение 5 минут.

Следующим этапом является формирование новой маски из фоторезиста методом оптической фотолитографии, аналогично основному электроду, для определения геометрии изоляции формируемого НИС перехода. Перед каждым нанесением фоторезиста образец проходит очистку в органическом растворителе подобную той, что проходят подложки. В установке Kurt J. Lesker проводится чистка образца в аргоновой плазме в течение 2 минут при давлении газа $6 \cdot 10^{-3}$ мбар, затем производится DC магнетронное напыление слоя изоляции (SiO_2) при мощности ~ 250 Вт и давлении аргона $6 \cdot 10^{-3}$ мбар. Затем лишний фоторезист удаляется методом взрывной литографии так же, как и на этапе формирования базового электрода. В результате этих действий формируется структуры «колодцев» различного диаметра, в которых на следующем этапе будут сформированы НИС переходы.

На третьем этапе происходит непосредственное формирование НИС перехода. Аналогично предыдущим этапам формируется маска для задания геометрии верхнего алюминиевого (С) электрода. Перед процессом напыления в установке Kurt J. Lesker производится травление в плазме аргона (ионное травление) в течении 2 минут при мощности 200 Вт, что позволяет удалить оксид меди на поверхности образца, который образуется вследствие разрыва вакуума. Далее происходит магнетронное DC напыление тонкого 7 нм слоя алюминия при давлении аргона $6 \cdot 10^{-3}$ мбар и мощностью 250 Вт на магнетроне. После этого для формирования слоя изолятора в НИС структуре в камеру напускается чистый кислород до достижения давления 10 мбар, алюминий в такой среде за 30 минут успевает достаточно глубоко окислиться с образованием равномерной оксидной пленкой вглубь от его поверхности, таким образом получается изоляционный слой AlO_x (И) толщиной ~ 1.5 нм, при этом влияние оставшегося чистого алюминия на туннельный переход за счет эффекта близости нивелируется. После этого сразу же проводится магнетронное DC напыление верхнего более толстого алюминиевого (С) контакта в атмосфере аргона при мощности 600 Вт, поданной на магнетрон, и давлении газа $3.6 \cdot 10^{-3}$ мбар. В конце производится финальное удаление фоторезиста с лишним металлом на нём методом «взрывной» литографии.

Для защиты верхних алюминиевых контактных площадок от окисления на воздухе на них может наноситься тонкий слой золота тем же магнетронным напылением.

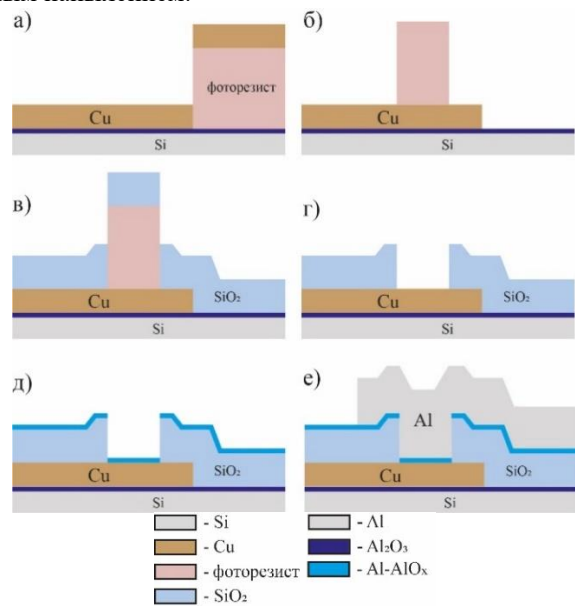


Рис.2. Порядок действий при изготовлении НИС переходов гибридных структур

Микроскопический анализ полученных гетероструктур

В первую очередь изготовленные образцы проходили микроскопический анализ в оптическом микроскопе, заметных недостатков обнаружено не было, после чего были проведены измерения с использованием оптического профилометра (рис.3), основанного на конфокальной технологии микроскопа ContourGT-3d. Далее, для более детального исследования, образцы анализировались в электронном микроскопе модели RAITH.

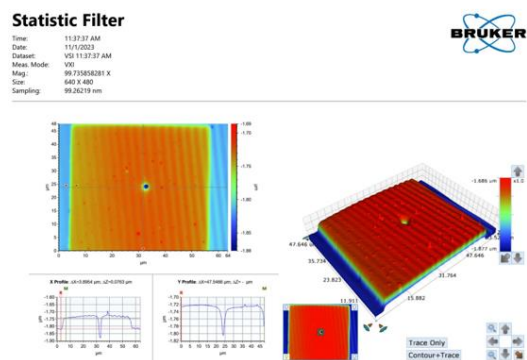


Рис.3. Изображение и профиль образца TA_4 с переходом 4 мкм в центре, полученные в оптическом профилометре

Более наглядное представление этой же структуры достигается при использовании электронного микроскопа с высоким разрешением и наблюдении под углом к поверхности, как показано на рис.4.

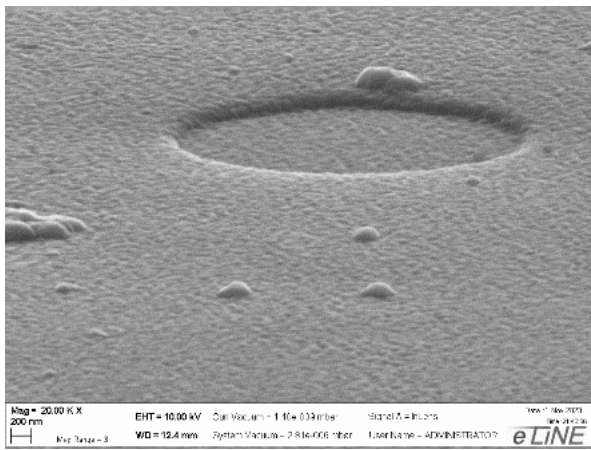


Рис. 4. Изображение перехода №10, полученное в электронном микроскопе при наблюдении под углом. Диаметр перехода 4 мкм. Площадь перехода 12 мкм²

Параметры изготовленных образцов

В рамках работы было изготовлена серия из четырех образцов ТА_1, ТА_2, ТА_3, ТА_4 с различными параметрами: диаметр и площадь туннельного перехода (1.5-5.2 мкм), толщина сверхпроводящего (100, 180, 220 нм) и нормального контакта (30 и 100 нм), высота изолирующих «колодцев», в которых формируются НИС переходы (70, 130, 160 нм). Помимо этого для части образцов были пропущены процедуры травления в плазме аргона и чистки меди в кислороде для оценки их влияния на качество переходов. Сопротивления туннельных переходов измерялись при температурах 300 К и 77 К: при этом произведение нормального сопротивления на площадь перехода варьировалось от 1.5 до 3 кОм на квадратный микрон.

Заключение

По результатам измерений образцы, изготовленные по новой технологии без использования метода теневой литографии, обладают визуально качественными НИС переходами Al-AlO_x-Cu и приемлемым нормальными сопротивлением.

Для дальнейшего использования в условиях технологической базы ИРЭ РАН рассмотренная технология формирования НИС переходов Al-AlO_x-Cu будет оптимизирована, планируется изготовление образцов по этой же технологии, но с использованием термического напыления алюминия из лодочки для улучшения точности напыления изоляционного слоя в переходе.

Работа выполнялась при поддержке российского научного фонда, проект № 23-72-00018.

Список литературы

1. Aharonov Y., Bohm D. Significance of Electromagnetic Potentials in the Quantum Theory //The Physical Review. – 1959. – Vol. 115. – No. 3. – P. 485.
2. Arutyunov K. Yu., Auraneva N. P., Vasenko A. S. Spatially resolved measurement of nonequilibrium quasiparticle relaxation in superconducting Al //Physical review. B. – 2011. – Vol 83. – No 10. – P 104509.
3. Arutyunov, K. Yu., Chernyaev, S. A., Karabassov, T., Lvov, D. S., Stolyarov, V. S., Vasenko, A. S. Relaxa-

tion of Nonequilibrium Quasiparticles in Mesoscopic Size Superconductors //Journal of Physics: Condensed Matter. – 2018. – Vol. 30. – P. 343001.

4. Гурский А. С., Шаповалов Д. Л., Арутюнов К. Ю. Транспорт неравновесных квазичастичных возбуждений в сверхпроводящем алюминии //Известия РАН. Серия физическая. –2024. – № 2. – С. 8.

МОДЕЛИРОВАНИЕ СВЕРХПРОВОДЯЩИХ ТОНКИХ ПЛЕНОК С ИСПОЛЬЗОВАНИЕМ НЕСТАЦИОНАРНЫХ УРАВНЕНИЙ ГИНЗБУРГА-ЛАНДАУ С ПОМОЩЬЮ ПЕРЕМЕННЫХ СВЯЗКИ

Паиковская В.Д.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В последние годы некоторые из наиболее важных тем физики опираются на сверхбыстрые однофотонные детекторы: квантовая запутанность и проблема затухания волновой функции во времени; обнаружение фотонов в диапазоне частот темной энергии и т. д. Этот класс электронных устройств может принести несколько прорывов в будущие научные исследования. Идея состоит в том, чтобы исследовать импульсы напряжения в небольших сверхпроводящих образцах вблизи границы между системами типов I и II (интертипный режим). Некоторые из наиболее многообещающих открытий, сделанных в последнее время в области сверхпроводимости, связаны с системами в области информационных технологий. Ранее теоретически описано, что стационарные состояния в сверхпроводниках в интертипном домене (или режиме) имеют очень похожие энергии, и это означает, что подвижность магнитного потока внутри сверхпроводников становится менее дорогостоящей, что требует исследования этого при различных механизмах рассеивания.

Модель

Обобщенные нестационарные уравнения Гинзбурга-Ландау в безразмерном виде имеют вид:

$$\frac{u}{\sqrt{1 + \gamma^2 |\Psi|^2}} \left(\frac{\partial}{\partial t} + i\varphi + \frac{1}{2} \gamma^2 \frac{\partial}{\partial t} |\Psi|^2 \right) \Psi = (\nabla - iA)^2 \Psi + (1 - T - |\Psi|^2) \Psi$$

и

$$\left(\frac{\partial A}{\partial t} + \nabla \varphi \right) = J_s - \kappa^2 \nabla \times \nabla \times A,$$

где плотность сверхпроводящего тока равна $J_s = \text{Im}[\Psi^* (\nabla - iA) \Psi]$

Используя теорему Гельмгольца, можно провести более простое тепловое моделирование, в результате чего получится уравнение теплопроводности:

$$C'_{eff} \frac{\partial T}{\partial t} = K_{eff} \nabla^2 T + \frac{1}{2} W_{Total} - \eta(T - T_0),$$

где T_0 – температура резервуара,

$C'_{eff} = \pi^4/48u$ – эффективная теплоемкость,

$K_{eff} = \pi^4/48u^2$ – эффективная теплопроводность.

Для определения общей диссипации необходимо обобщить теорему о свободной энергии Гельмгольца, выраженную в безразмерных единицах:

$$W_{Total} = 2 \left(\left| \frac{\partial A}{\partial t} \right|^2 \right) + \frac{2u}{\sqrt{1 + \gamma^2 |\Psi|^2}} \left[\left| \frac{\partial \Psi}{\partial t} \right|^2 + \frac{\gamma^2}{4} \left(\frac{\partial}{\partial t} |\Psi|^2 \right)^2 \right]$$

Граничные условия системы имеют вид:

$$\begin{aligned} \hat{n} \cdot (\nabla - A)\Psi &= 0 \\ B|_{\partial V} &= H_e \\ T|_{\partial V} &= T_0 \end{aligned}$$

Результаты

Обнаружена эволюция параметра порядка, магнитного поля и локальной температуры при значениях κ , близких к $\kappa = 1/\sqrt{2}$ и $T = T_c$. В системе граница влияет на проникновение вихрей и их устойчивую конфигурацию. Полученные результаты имеют несколько ограничений, поскольку симметрия была аппроксимирована в направлении оси OZ. В таких случаях поле рассеяния может внести важный вклад в магнитное поле.

На рис.1 показана полная теплота, рассеиваемая по всему образцу, в зависимости от времени при всех значениях $\kappa = 0,5; 0,6; 0,7; 0,8$. Наблюдаемые пики соответствуют вхождению вихрей в образец. Второй экстремум на графике соответствует повторному входу новой партии вихрей в образец. Можно заметить, что локальный отвод тепла увеличивает локальную температуру, что вызывает снижение значений параметра порядка, что облегчает проникновение вихрей.

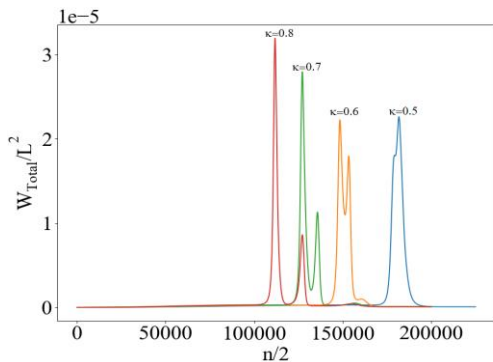


Рис.1. Временное изменение общей диссипации

Данные результаты показывают стандартные вихри, которые не должны существовать в стандартной Абрикосовской картине при $\kappa < 1/\sqrt{2}$. Из-за тепловой диссипации барьер Бина-Ливингстона ослабевает, что способствует проникновению вихрей. На рис.2 и рис.3 показаны изображения временной эволюции параметра порядка и температуры. В процессе движения вихрей можно наблюдать рост локальной температуры внутри образца, когда как на границе температура остается неизменной ($\Delta T = 0$) ввиду граничных условий (мы полагаем, что термализация с термальной баней мгновенная). В момент проникновения 4-х вихрей в образец возникает изменение в магнитном потоке, из-за чего возникает напряжение в образце (закон Фарадея). Одним из вкладов в общую диссипацию является $W_A = 2 \left(\left| \frac{\partial A}{\partial t} \right|^2 \right)$, что означает что движение магнитного потока генерирует рост температуры по линии проникновения вихрей.

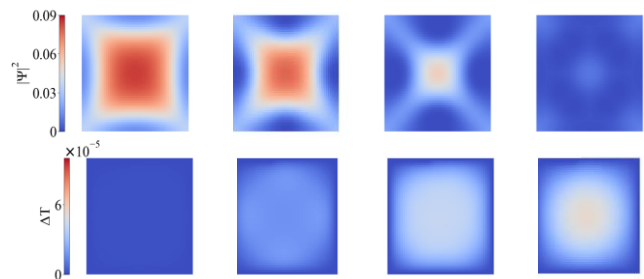


Рис.2. Эволюция параметра порядка и температуры во времени ($\kappa = 0,5$) для различных значений внешнего поля 0,175, 0,178, 0,181, 0,184 ($Hc2(0)$)

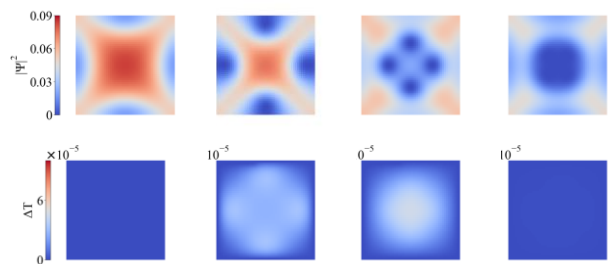


Рис.3. Эволюция параметра порядка и температуры во времени ($\kappa = 0,8$) для различных значений внешнего поля 0,1075, 0,111, 0,114, 0,12 ($Hc2(0)$)

Заключение

Представлено численное интегрирование обобщенных уравнений динамики Гинзбурга-Ландау. Для рассмотрения была выбрана двумерная система с перпендикулярным ей магнитным полем. За счет увеличения поля было достигнуто поле, при котором вихри могут проникать в образец. Полученные результаты надежно коррелируют с предыдущими результатами для аналогичных систем: величина магнитного поля уменьшается с увеличением параметра Гинзбурга-Ландау, пересекающего границу между

сверхпроводимостью 1-го и 2-го рода. Установлено, что даже при наличии вихрей они проникают и в образец. Этот результат верен, так как эта величина была определена для бесконечной системы, в то время как в данной работе мы рассматриваем мезоскопическую систему с квадратной стороной, равной $24\xi(0)$ приложенным магнитным полем, увеличивающимся с небольшой, но постоянной скоростью, создавая постоянную диссипацию. Важно отметить, что при приближении к точке Богомольного мы также наблюдаем нормальные значения теплоотдачи.

Благодарность

Авторы благодарят программу сотрудничества «Зеркальные лаборатории» Национального Исследовательского Университета Высшая Школа Экономики.

Список литературы

1. Gol'tsman G.N. et al. Picosecond superconducting single-photon optical detector // *Appl. Phys. Lett.* 2001. Vol. 79, № 6. P. 705–707.
2. Vagov A. et al. Superconductivity between standard types: Multiband versus single-band materials // *Phys. Rev. B.* 2016. Vol. 93, № 17. P. 174503.
3. Duarte E.C.S. et al. Comparing energy dissipation mechanisms within the vortex dynamics of gap and gapless nano-sized superconductors. // *Materials Science and Engineering: B.* 2023. Vol. 296, P. 1

ИССЛЕДОВАНИЕ НЕРАВНОВЕСНЫХ
ПРОЦЕССОВ В ТУННЕЛЬНЫХ КОНТАКТАХ
МЕТАЛЛ - СВЕРХПРОВОДНИК

Амиров Э.Ш.¹, Васенко А.С.²

Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент прикладной математики,

²департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В нашей работе изучено влияние неравновесных процессов на квазичастичный электрический ток и на функции распределения в туннельных контактах нормальный металл (N) – изолятор (I) – неравновесный сверхпроводник (S') – сверхпроводник (S). Построены графики квазичастичного тока и функций распределения при различных температурах с учетом неупругой релаксации частиц в неравновесном сверхпроводнике S'. Показано, что при отсутствии сильной релаксации электрический ток для напряжений $eV < \Delta$ подавлен. Мы связываем это подавление с обратным туннелированием квазичастиц в нормальный металл.

Введение

Сверхпроводниковые гибридные структуры на протяжении последних десятилетий привлекают серьезный научный интерес. В их составе кроме сверхпроводника могут быть нормальные металлы и ряд других материалов (ферромагнетики, ферромагнитные изоляторы, графен, топологические изоляторы и т.д.). В данное время сверхпроводниковые гибридные структуры нашли применение в устройствах классической и квантовой логики [1], высокоточных сенсорах электромагнитных сигналов [1 – 4], в качестве переменных резисторов [3] и в качестве микро и нанохолодильников [1 – 5].

Механизмы переноса тока из нормального металла в сверхпроводник и сопутствующего переноса тепла хорошо известны и описаны в обзоре [1]. Данные процессы связаны с наличием в сверхпроводнике энергетической щели Δ , которая вызывает избирательное туннелирование квазичастиц с энергией $E > \Delta$ из нормального металла в сверхпроводник. Такой ток называют одночастичным, и именно с ним связано электронное охлаждение в нормальном металле при подщелевых напряжениях [6, 7]. При $eV \lesssim \Delta$ одночастичный ток экспоненциально спадает с понижением температуры. При очень низких температурах в подщелевой области ток осуществляется посредством двухчастичного Андреевского отражения для частиц с энергией $E < \Delta$ [8, 9]. При этом два электрона из нормального металла туннелируют в сверхпроводник, образуя куперовскую пару. В результате при достаточно сильном сопротив-

лении туннельного контакта андреевский ток можно считать пренебрежимо малым, если не рассматривать сверхнизкие температуры.

При таких условиях главным лимитирующим фактором устройств на базе NIS контактов является неравновесная инжекция квазичастиц из нормального металла в сверхпроводник. В результате сверхпроводник может перегреваться, что приводит к подавлению энергетической щели, а также возможно обратное туннелирование неравновесных квазичастиц в нормальный металл.

В данной работе изучается неравновесная инжекция квазичастиц с последующей неупругой релаксацией в сверхпроводнике. С этой целью была рассмотрена NIS'S структура, описанная в следующей секции. Результаты, полученные нами, могут быть обобщены для широкого класса сверхпроводниковых гибридных структур.

Модель и уравнения

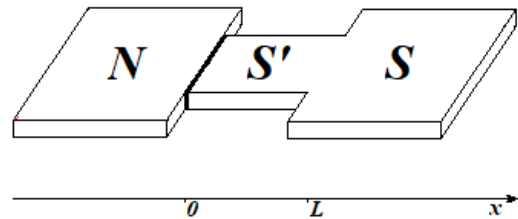


Рис.1. Геометрия рассматриваемой системы

На рис.1 изображена NIS'S структура, которая состоит из нормально металлического резервуара (N), изолятора (I), сверхпроводникового провода S' толщины L и сверхпроводникового резервуара S. Провода S и S' сделаны из одного и того же материала (например, алюминия), и граница между ними предполагается абсолютно прозрачной. Рассматривается диффузионный предел, при котором длина упругого рассеяния ℓ намного меньше длины когерентности $\xi_0 = \sqrt{D/2\Delta}$, где D – коэффициент диффузии (предполагается, что $\hbar = k_B = 1$). Длина L много больше ξ_0 .

Вычисление тока требует решения одномерных уравнений Келдыша-Узаделя [11]

$$[\partial_x E + \check{\Delta}, \check{G}] = iD \partial \check{J}, \quad \check{J} = \check{G} \partial \check{G}, \quad \check{G}^2 = \check{1} \quad (1)$$

$$\check{G} = \begin{pmatrix} \hat{g}^R & \hat{G}^K \\ 0 & \hat{g}^A \end{pmatrix}, \quad \hat{G}^K = \hat{g}^R \hat{f} - \hat{f} \hat{g}^A, \quad (2)$$

где $\hat{g}^{R,A}$ – гриновские матричные функции 2×2 в пространстве матриц Намбу, $\hat{f} = f_+ + \sigma_z f_-$ – матричные функции распределения, $\sigma_{y,z}$ – матрицы Паули в пространстве Намбу, $\hat{\Delta} = e^{i\sigma_z \chi} i \sigma_y \Delta$, Δ и χ являются модулем и фазой парного потенциала,

$\partial \equiv \partial / \partial x$, $\mathbb{1}$ – единичная матрица, $\check{\sigma}_z$ и $\check{\Delta}$ – блочные матрицы с матрицами σ_z и $\hat{\Delta}$ на диагонали.

Граничные условия для функции \check{G} и матрицы спектрального тока \check{J} при $x = -0$ и $x = +0$ (слева и справа от туннельного барьера при $x = 0$) задаются соотношением [12]

$$(\sigma_N/g_N)\check{J}_{-0} = \check{J}_{+0}(W/\xi_0) = [\check{G}_{-0}, \check{G}_{+0}], \quad (3)$$

где σ_N и g_N проводимость нормального металла N и сверхпроводящего провода S' в нормальном состоянии, соответственно, а $W \ll 1$ – параметр прозрачности,

$$W = \frac{R(\xi_0)}{2R},$$

где $R(\xi_0) = \xi_0/g_N$ является сопротивлением провода S' в нормальном состоянии на длине ξ_0 , а R – сопротивление туннельного барьера. На прозрачной границе $S-S$ все функции и их первые производные непрерывны.

Электрический ток связан с келдышевской компонентой матрицы спектрального тока \check{J} как [13-16]

$$I = -\frac{g_N}{4e} \int_0^\infty \text{Tr} \sigma_z \check{J}^K dE, \quad (4)$$

и с учетом граничного условия (3)

$$I = \frac{1}{8eR} \int_0^\infty \text{Tr} \sigma_z [\check{G}_{-0}, \check{G}_{+0}]^K dE. \quad (5)$$

Поскольку мы пренебрегаем двухчастичными процессами андреевского отражения, а следовательно эффектом близости, то можно считать, что функции Грина в сверхпроводниковом проводе S' совпадают с их равновесными значениями в резервуаре S [11]. При этом функции распределения $f_{\pm S}(x, E)$ должны быть найдены из кинетических уравнений, которые получаются из келдышевской компоненты уравнения (1), и для $E > 0$ в нашем случае имеют вид $\partial^2 f_{+S}(x, E) = 0$ и $\partial^2 f_{-S}(x, E) = 0$. Однако, обе функции распределения $f_{\pm S}(x, E)$ растут пропорционально x . Такой рост на практике ограничен неупругими столкновениями, благодаря которым обе функции релаксируют к равновесным значениям при $x \sim l_{\pm} \gg \xi_0$, где $l_{\pm} = \sqrt{\mathcal{D}\tau_{\pm}}$ – длины неупругого рассеяния, а τ_{+} и τ_{-} – времена неупругого рассеяния. Для простоты вместо включения интегралов столкновения в кинетические уравнения мы используем τ -приближение,

$$l_{+}^2 \partial^2 f_{+S} = (f_{+S} - f_{eq})N(E), \quad (6)$$

$$l_{-}^2 \partial^2 f_{-S} = f_{-S}N(E), \quad (7)$$

где $N(E) = \theta(|E| - \Delta)|E|/\sqrt{E^2 - \Delta^2}$ – плотность состояний в сверхпроводнике.

Граничные условия при $x = 0$ для уравнений (6) и (7) получаются из выражения (3). На правой прозрачной границе функции распределения становятся равновесными функциями сверхпроводящего резервуара S

$$f_{+S} |_{x=L} = f_{eq} = \tanh\left(\frac{E}{2T_S}\right), \quad f_{-S} |_{x=L} = 0. \quad (8)$$

При расчетах, для учета зависимости значения энергетической щели от температуры была использована формула из книги [17].

Неравновесное распределение в S' и одночастичный ток

Решение уравнения (7) с начальными условиями (3) и (8) позволяет получить $f_{-S}(x, E)$ в S' , которое описывает релаксацию дисбаланса дырок и электронов

$$f_{-S} = \frac{f_{-N}\alpha_{-}}{\left(\alpha_{-} + \sqrt{N(E)} \coth\left(\beta_{-}/\sqrt{N(E)}\right)\right) \sinh\left(\left(\beta_{-} - x/l_{-}\right)/\sqrt{N(E)}\right)}, \quad (9)$$

где $\alpha_{-} = 2Wl_{-}/\xi_0$, $\beta_{-} = L/l_{-}$, $f_{-N} = \frac{1}{2}\left(\tanh\left(\frac{E+eV}{2T_N}\right) - \tanh\left(\frac{E-eV}{2T_N}\right)\right)$. Данная функция экспоненциально спадает с характерной длиной $x \sim l_{-}\sqrt{N(E)}$.

Из (6) с учетом (3) и (8) получаем также $f_{+S}(x, E)$ в S' ,

$$f_{+S} = f_{eq} - \frac{(f_{eq} - f_{+N})\alpha_{+}\sqrt{N(E)}}{\alpha_{+}\sqrt{N(E)} + \coth\left(\beta_{+}\sqrt{N(E)}\right) \sinh\left(\left(\beta_{+} - x/l_{+}\right)\sqrt{N(E)}\right)}, \quad (10)$$

где $\alpha_{+} = 2Wl_{+}/\xi_0$, $\beta_{+} = L/l_{+}$, $f_{+N} = \frac{1}{2}\left(\tanh\left(\frac{E+eV}{2T_N}\right) + \tanh\left(\frac{E-eV}{2T_N}\right)\right)$. Следует отметить, что функция $f_{+N}(x, E)$ не влияет на значение тока, но влияет на энергетический ток в NIS структуре.

На рис.2 и рис.3 изображены функции $f_{-S}(E, x)$ и $\delta f_{+S}(E, x) = f_{eq} - f_{+S}$ при температурах $0.3T_C$ и $0.2T_C$. На графиках видно, что вклад неравновесных эффектов максимален при энергиях близких к Δ . Для $E \gg \Delta$ обе функции стремятся к нулю.

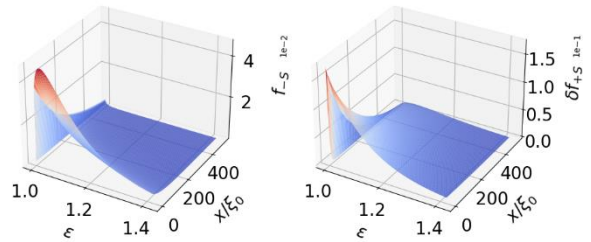


Рис.2. Слева $f_{-S}(E, x)$ и справа $\delta f_{+S}(E, x)$. $T_N = T_S = 0.3T_C$, $eV = 0.8\Delta$, $L = 2000\xi_0$, $l_{-} = 50$, $l_{+} = 400\xi_0$, $W = 0.01$, $\varepsilon = E/\Delta$

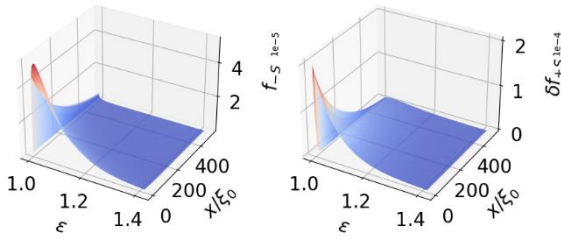


Рис.3. Слева $f_{-S}(E, x)$ и справа $\delta f_{+S}(E, x)$. $T_N = T_S = 0.2T_C$, $eV = 0.8\Delta$, $L = 2000\xi_0$, $l_- = 50$, $l_+ = 400\xi_0$, $W = 0.01$, $\varepsilon = E/\Delta$

Решив уравнение (7) с учетом условий (3) и (8), а затем воспользовавшись выражением (5), получим выражение для одночастичного тока

$$I = \frac{1}{eR} \int_{\Delta}^{\infty} \frac{N(E)f_{-N} \coth(\beta_-/\sqrt{N(E)})}{\alpha_-/\sqrt{N(E)} + \coth(\beta_-/\sqrt{N(E)})} dE. \quad (11)$$

На рис.4 и рис.5 изображены вольт-амперные характеристики NIS'S контактов при разных температурах ($0.2T_C$ и $0.3T_C$). На обоих графиках видно, что с ростом l_- (длины неупругой релаксации зарядового дисбаланса [17]) ток уменьшается. Когда $l_- > L$ ток почти не зависит от l_- и L (длина S') играет роль длины релаксации.

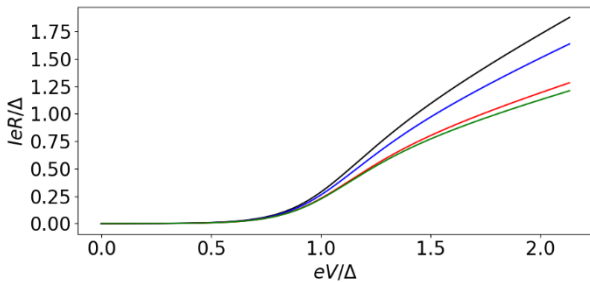


Рис.4. Вольт-амперная характеристика NIS'S структуры. $T_N = T_S = 0.3T_C$, $W = 0.01$, $L = 50\xi_0$. Черным цветом изображен равновесный ток. Синим, красным, зеленым одночастичный ток при $l_- = 10\xi_0, 50\xi_0, 300\xi_0$ соответственно

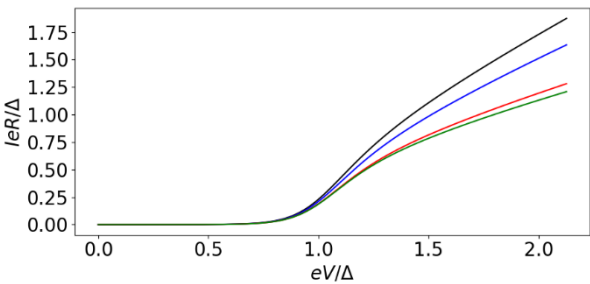


Рис.5. Вольт-амперная характеристика NIS'S структуры. $T_N = T_S = 0.2T_C$, $W = 0.01$, $L = 50\xi_0$. Черным цветом изображен равновесный ток. Синим, красным, зеленым одночастичный ток при $l_- = 10\xi_0, 50\xi_0, 300\xi_0$ соответственно

Заключение

Было исследовано влияние неравновесной инжекции квазичастиц на электрический ток и функции распределения в сверхпроводниковом проводе в NIS'S контакте при разных длинах неупругой релаксации. Была построена модель для описания неравновесной квазичастичной инжекции.

Было показано, что когда длина релаксации в сверхпроводнике достаточно велика по сравнению с длиной когерентности, ток при $eV < \Delta$ подавлен. Мы связываем это подавление с обратным туннелированием квазичастиц в нормальный металл.

Список литературы

1. Giazotto F. Heikkilä T. T., Luukanen A., Savin A. M., Pekola J. P. Opportunities for mesoscopics in thermometry and refrigeration: Physics and applications [Journal] // Reviews of Modern Physics. - 2006. - 1 : Vol. 78. - pp. 217-274.
2. Nahum M. Eiles T. M., Martinis J. M. Electronic microrefrigerator based on a normal-insulator-superconductor tunnel junction [Journal]. - 1994. - 24 : Vol. 65. - pp. 3123-3125.
3. N. Ullom J. Physics and applications of NIS junctions. - American Institute of Physics [Conference] // Physics and applications of NIS junctions. - 2002. - Vol. 605. - pp. 135-140.
4. Clark A. M. Miller N. A., Williams A., Ruggiero S. T., Hilton G. C., Vale L. R., Beall J. A., Irwin K. D., Ullom J. N. Cooling of bulk material by electron-tunneling refrigerators [Journal] // Applied Physics Letters. - 2005. - 17 : Vol. 86. - pp. 1-3.
5. Muhonen J. T. Meschke M., Pekola J. P. Micrometre-scale refrigerators [Journal] // Reports on Progress in Physics. - 2012. - 4 : Vol. 75. - p. 046501.
6. Vasenko A. S. Bezuglyi E. V., Courtois H., Hekking F. W.J. Electron cooling by diffusive normal metal-superconductor tunnel junctions [Journal] // Physical Review B - Condensed Matter and Materials Physics. - 2010. - 9 : Vol. 21. - pp. 221-232.
7. Vasenko A. S. Hekking F. W. J. Nonequilibrium electron cooling by NIS tunnel junctions [Journal] // Journal of Low Temperature Physics. - 2009. - 5-6 : Vol. 154. - pp. 221-232.
8. Andreev A. F. The thermal conductivity of the intermediate state in superconductors [Journal] // JETP. - 1964. - 5 : Vol. 19. - pp. 1823-1828.
9. Rajauria S. Gandit P., Hekking F. W. J., Pannetier B., Courtois H. Competition between electronic cooling and Andreev dissipation in a superconducting micro-cooler [Journal] // Journal of Low Temperature Physics. - 2009. - 154. - pp. 211-220.
10. Pekola J. P. Anghel D. V., Suppala T. I., Suoknuuti J. K., Manninen A. J., Manninen M. Trapping of quasiparticles of a nonequilibrium superconductor [Journal] // Applied Physics Letters. - 2000. - 19 : Vol. 76. - pp. 2782-2784.
11. Belzig W. Wilhelm F. K., Bruder C., Schön G., Zaikin A. D. Quasiclassical Green's function approach to mesoscopic superconductivity [Journal] // Superlattices and Microstructures. - 1999. - 5-6 : Vol. 25. - pp. 1251-1288.

12. Куприянов М. Лукичев В. Влияние прозрачности границ на критический ток грязных SS'S структур // ЖЭТФ. - 1988 г. - 139 : Т. 94.

13. Bardas A. Averin D. Peltier effect in normal metal -- superconductor microcontacts [Journal] // Physical Review B. - 1995. - 17 : Vol. 52. - pp. 12873-12877.

14. Virtanen P. Heikkilä T. Thermoelectric effects in superconducting proximity structures [Journal] // Applied Physics A. - 2007. - Vol. 89. - pp. 625-637.

15. Bezuglyi E. V. Vinokur V. Heat transport in proximity structures [Journal] // Physical review letters. - 2003. - 13 : Vol. 91. - p. 137002.

16. Yokoyama T. Tanaka Y., Golubov A. A., Asano Y. Theory of thermal and charge transport in diffusive normal metal/superconductor junctions [Journal] // Physical Review B. - 2005. - 21 : Vol. 72. - p. 214513.

17. Tinkham M Introduction to superconductivity [Book]. - [s.l.] : Courier Corporation, 2004. - p. 63.

МОДЕЛИРОВАНИЕ НАНОПРОВОЛОК С ИСПОЛЬЗОВАНИЕМ НЕСТАЦИОНАРНЫХ УРАВНЕНИЙ ГИНЗБУРГА-ЛАНДАУ

Кононова Н.А.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Данная работа посвящена использованию нестационарных уравнений Гинзбурга-Ландау для численного моделирования двух рядом стоящих сверхпроводящих проводов и прогнозирования характеристик их намагниченности во внешнем магнитном поле в зависимости от температуры. В работе приведено физическое описание исследуемых явлений, алгоритм работы программы, а также продемонстрированы результаты моделирования.

Введение

Начиная с открытия сверхпроводимости в 1911 году, ведутся интенсивные исследования физики сверхпроводников и их применения. Одной из важных характеристик сверхпроводников является их поведение в приложенном магнитном поле, чему посвящено большое количество экспериментальных и теоретических работ. Сверхпроводящие материалы сегодня имеют широкий спектр применений. Они могут быть использованы, например, в областях энергетики, медицинской диагностики, при производстве сверхпроводящих квантовых интерферометров, которые, в свою очередь, могут быть использованы в качестве высокочувствительных детекторов слабых магнитных полей.

Современные методы производства позволяют создавать сверхпроводящие образцы различных размеров, включая наномасштабные структуры. Массивы сверхпроводящих нанопроводов могут быть по-

лезны в качестве элементов новых нанoeлектронных устройств. Сверхпроводящие нанопроволоки уже применяются в различных областях, включая детекторы одиночных фотонов, квантовые схемы и квантовые компьютеры. Для исследования и анализа свойств этих мезоскопических сверхпроводников широко применяется численное интегрирование нестационарных уравнений Гинзбурга-Ландау (ГЛ). Этот метод позволяет получать данные о параметрах порядка и магнитном поле, которые являются важными характеристиками для описания поведения подобных систем.

В статье А.А. Нояна и др. [1] были измерены значения намагниченности M образцов сверхпроводящих нанопроволок в зависимости от магнитного поля H и температуры T в продольном и поперечном поле. Результаты эксперимента представлены на рис. 1. Диаметр нанопроволок In имел порядок величины 200 нм, сравнимый со значением длины когерентности, благодаря чему было возможно изучение геометрически ограниченных сверхпроводящих состояний.

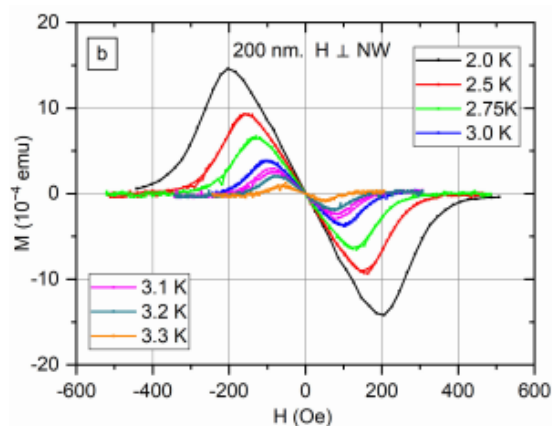


Рис.1. Магнитный момент M как функция поперечного магнитного поля H для массивов нанопроволок In с диаметром 200 нм [1]

Описание моделирования

В этой работе моделирование для решения нестационарных уравнений Гинзбурга-Ландау по методу Эйлера выполнялось таким образом, что магнитное поле изменялось с фиксированным шагом при постоянной температуре.

Применяются комплексные переменные связки [3] U_x , U_y и U_z для сохранения калибровочно-инвариантных свойств дискретизированных уравнений.

Вся конфигурация имеет форму прямоугольного ящика, занимающего ограниченную область в пространстве. Проволоки располагаются в декартовой системе координат, параллельно оси y . Размеры системы по осям x , y , z - 108, 320, 88 в единицах ξ , где ξ - длина когерентности. Магнитное поле задано однородно, по направлению оси z . Размер сетки, используемой при дискретизации для вычислений - 0.125ξ . На рис.2 приводятся размеры и положение нанопроволок:

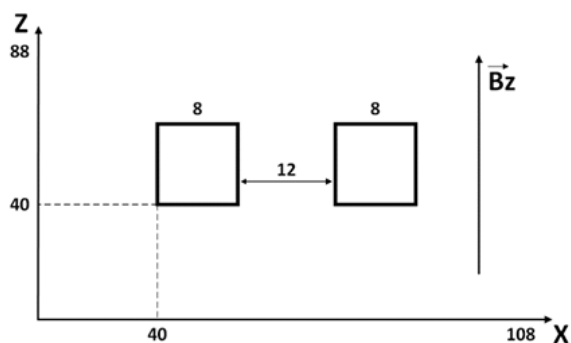


Рис.2. Параметры моделирования

Теория сверхпроводимости Гинзбурга — Ландау вводит комплексную псевдоволновую функцию Ψ как параметр порядка для сверхпроводящих электронов, таким образом, что для сверхпроводящих электронов локальная плотность равняется квадрату ее модуля [2].

Кроме того, можно считать, что волновая функция Ψ описывает движение центра масс куперовских пар. Рассматриваемая программа базируется на вычислении величины Ψ в каждой дискретной точке системы для каждого дискретного значения магнитного поля.

Моделирование проводится с помощью нестационарных уравнений Гинзбурга-Ландау (1) и (2) [3]:

$$\frac{\partial \Psi}{\partial t} = - \frac{\partial F}{\partial \Psi^*} \quad (1)$$

$$\sigma \frac{\partial \vec{A}}{\partial t} = - \frac{1}{2} \frac{\partial F}{\partial \vec{A}} - \vec{\nabla} \times \vec{\nabla} \times \vec{A} \quad (2)$$

где Ψ - комплексное значение параметра порядка, \vec{A} - вектор-потенциал, F - свободная энергия Гинзбурга-Ландау.

Можно выделить следующие условия применимости нестационарных уравнений ГЛ: требование, чтобы функции Ψ и A изменялись в пространстве достаточно медленно, а температура не слишком сильно отличалась от критической температуры T_c [4]. В данной работе моделирование проводилось при дробных значениях критической температуры: 0.6 T_c , 0.7 T_c , 0.8 T_c , 0.9 T_c , иначе называемой температурой перехода в сверхпроводящее состояние.

Параметр теории ГЛ определяется отношением (3) длины когерентности ξ и глубины проникновения магнитного поля λ :

$$\kappa = \frac{\lambda}{\xi} \quad (3)$$

Для типичных чистых сверхпроводников $\lambda = 500 \text{ \AA}$ и $\xi = 3000 \text{ \AA}$, так что $\kappa \ll 1$. В нашем случае, параметры моделирования заданы так, что $\kappa = 0.25$, в связи с тем, что в проведенном эксперименте [1] значение длины когерентности равняется $\xi = 264 \text{ нм}$, а средняя глубина проникновения равна $\lambda = 53 \text{ нм}$.

В зависимости от величины κ сверхпроводники подразделяются на сверхпроводники 1 и 2 рода. Для

сверхпроводников 1 рода характерен резкий переход в сверхпроводящее состояние при одном фиксированном значении поля. Магнитное поле из объема сверхпроводника вытесняется полностью. Величина $\kappa = 0.25$ соответствует первому роду. Второй род сверхпроводников при переходе в сверхпроводящее состояние характеризуется двумя значениями поля H_1 и H_2 , между которыми при возникновении сверхпроводимости наблюдается смешанное состояние.

Магнитное поле H измеряется в единицах критического магнитного поля объемного сверхпроводника и изменяется от $-4H_c$ до $4H_c$. Провод считается бесконечно длинным.

Для первого значения H системе задано состояние, где $U_x(i, j, k) = U_y(i, j, k) = U_z(i, j, k) = 1$.

Алгоритм решения можно описать следующим образом. После задания начальных условий, стартует цикл изменения магнитного поля в заданном диапазоне $H+dH$, внутри которого запускается цикл $tn+dt$ по числу шагов решения, содержащий в себе уравнения Гинзбурга-Ландау, расчет граничных условий и проверку на критерий сходимости [5]. При фиксированной температуре, магнитное поле H увеличивается с небольшими интервалами ΔH . Стационарное состояние, найденное для фиксированного значения H , затем используется в качестве начального условия для $H + \Delta H$. Таким образом, исследуется эволюция системы во времени за счет полученных результатов для каждого фиксированного магнитного поля, приложенного к системе. После завершения вложенного цикла, в основном цикле также происходит расчет намагниченности, расчет величины магнитного поля B внутри конфигурации, расчет локальной плотности сверхпроводящих электронов.

Анализ результатов

В 1933 г. Мейснером и Оксенфельдом было обнаружено вытеснение магнитного поля из объема материала при его переходе в сверхпроводящее состояние [6], когда он охлаждается ниже определенной температуры T_c . Это явление носит название *эффект Мейснера-Оксенфельда*.

Существует также обратимый эффект Мейснера. Его суть в том, что сверхпроводящее состояние разрушается под воздействием критического магнитного поля H_c , которое связано с энергией конденсации сверхпроводящего состояния. Внешнее магнитное поле возвращает образец в его нормальное удельное сопротивление при достаточно высоких собственных значениях. Таким образом, сверхпроводник характеризуется двумя критическими параметрами: температурой и магнитным полем.

Результаты моделирования для значения $H = -1.6$ и температур 0.6 T_c , 0.7 T_c , 0.8 T_c , 0.9 T_c представлены на рис.3:

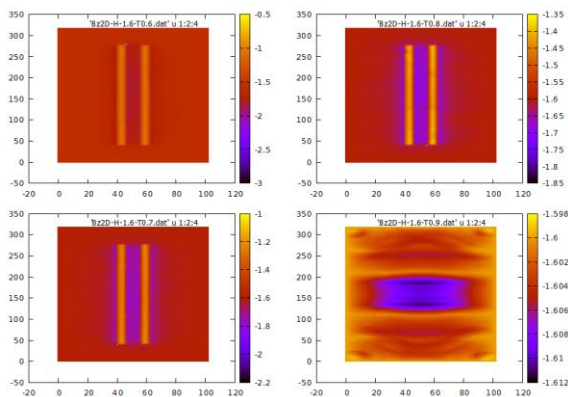


Рис.3. Магнитное поле H для сверхпроводящих проволок

Как можно видеть, в случае 0.9Тс сверхпроводимость разрушена и магнитное поле проникло в образец. Это обуславливается обратимым Мейснеровским эффектом: тем, что значение магнитного поля в данной точке $H = H_c$, критическому значению поля. Для более низких температур виден «классический» Мейснеровский эффект. Магнитное поле выталкивается сверхпроводником, незначительно проникая в него.

Результаты моделирования зависимости намагниченности M от величины магнитного поля H представлены на графике (рис.4). В данном моделировании расстояние между нанопроволоками равнялось $12 \cdot 0.125\xi$. Значение намагниченности усредняется относительно всех точек конфигурации:

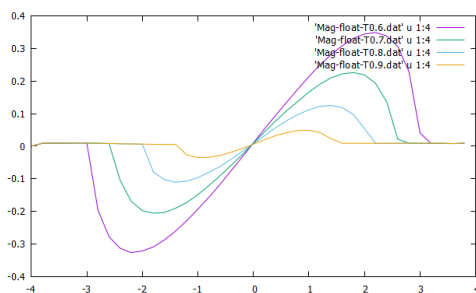


Рис.4. Зависимость намагниченности от величины магнитного поля H

Кроме того, очевидно, что для данного сверхпроводника первого рода наблюдается поведение, характерное для второго рода: частичное проникновение поля в нанопроволоки перед разрушением сверхпроводимости. Помимо этого, сверхпроводимость разрушается при значении магнитного поля, соответствующего 1.8-3.1Тс для объемного сверхпроводника, то есть сверхпроводимость в нанопроволоках существует выше критического магнитного поля. Это вызвано тем, что размер нанопроволоки слишком мал для того, чтобы образец вел себя в полном соответствии с разделением на рода сверхпроводников.

Увеличение расстояния между проволоками до $100 \cdot 0.125\xi$ качественно не изменило графики. Критические значения магнитного поля фактически не изменились. Стал чуть более выраженно заметен гистерезис. На рис.5 показан график при 0.8Тс:

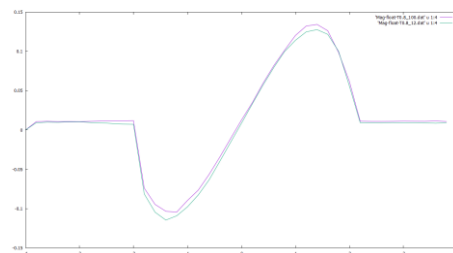


Рис.5. Зависимость намагниченности от величины магнитного поля H

Результаты моделирования качественно совпадают с приведенными в статье А. А. Нояна и др. [1].

Заключение

Таким образом, было проведено моделирование для двух сверхпроводящих проводов, находящихся в изменяющемся магнитном поле для разных температур. Мы проследили за временной эволюцией локального магнитного поля, получив набор стационарных решений.

В результате анализа полученных характеристик был сделан вывод, что с уменьшением температуры, увеличивается значение критического поля H_c , при котором разрушается сверхпроводимость. При увеличении расстояния между сверхпроводящими нанопроволоками, критическое поле не изменяется, возрастает только максимальное значение намагниченности для системы при данной температуре. Данная информация может быть использована при выборе плотности расположения нанопроволок при создании метаматериала.

Список литературы

1. Noyan A.A. Size-Dependent Superconducting Properties of in Nanowire Arrays / Noyan A.A., Ovchenkov Y.A., Ryazanov V.V., Golovchanskiy I.A., Stolyarov V.S., Levin E.E., Napolskii K.S. // *Nanomaterials* - 2022, 12, 4095. <https://doi.org/10.3390/nano12224095>
2. Тинкхам М. Введение в сверхпроводимость: Пер. с англ. — М.: Атомиздат, 1980.— Пер. изд.: США, 1975. 310 с.
3. D. Gropp et al. Numerical Simulation of Vortex Dynamics in Type-II Superconductors // *Journal of computational physics* – 1996, №123, p.254–266, article no. 002
4. Ружицкий В.И. Новые физические подходы к решению задачи масштабирования элементной базы цифровой сверхпроводниковой электроники : диссертация кандидата физико-математических наук : 01.04.15 / Ружицкий Всеволод Игоревич; [Место защиты: ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова»]. - Москва, 2021. - 121 с. : ил.
5. Camacho, Wilmer Yecid Córdoba. Vortex Matter in Intertype Superconductivity Regime: диссертация кандидата физико-математических наук – Recife, 2018
6. Meissner W., Ochsenfeld R. Ein neuer Effekt bei Eintritt der Supraleitfähigkeit - *Naturwissenschaften*, 1933, № 21, S. 787.

СВЧ-ГЕНЕРАТОР НА ОСНОВЕ ДЖОЗЕФСОНОВСКОГО ПЕРЕХОДА

Тюменев Р.¹, Калашиников Д.С.^{1,2},
Шишкин А.Г.^{1,2}, Столяров В.С.^{1,2}

¹Национальный исследовательский университет «МФТИ»,
физтех-школа физики и исследований
им. Ландау (ЛФИ),
²ФГУП ВНИИА им. Н.Л. Духова

Аннотация

В данной работе рассматривается генератор, представляющий собой планарную систему, состоящую из джозефсоновского перехода, микроволнового резонатора, шунтирующей ёмкости и шунтирующего сопротивления. Такой генератор работает при температуре 20 мК на частоте, соответствующей управлению кубитами.

Введение

Физические системы, используемые для квантовых вычислений, работающие в микроволновом диапазоне, требуют передовой управляющей электроники, а использование интегрированных компонентов, работающих при температуре квантовых устройств, потенциально выгодно.

Использование генератора на чипе внутри квантового компьютера позволяет упростить рабочую схему для управления кубитами. Такой генератор, работающий при температуре 20 мК на частоте, соответствующей управлению кубитами, уже был показан в работе [1], однако методика изготовления предоставленного чипа очень трудоемка. Целью работы является определение диапазона параметров генератора, в котором возможна стабильная генерация, путем численного решения уравнений динамики системы, изготовления отдельных элементов генератора, а также поиска его оптимальных параметров с помощью моделирования с учетом полученных элементов генератора.

Моделирование системы для поиска оптимальных параметров

Для определения диапазона рабочих параметров такой системы, при которых имеется возможность получения устойчивости генерации, было проведено моделирование. Моделирование основывается на численном решении уравнения (1), нормированного на критические величины джозефсоновского перехода, как более общего уравнения, описывающего динамику джозефсоновского перехода.

$$\beta\ddot{\varphi} + \dot{\varphi} + \sin(\varphi) = i_{ac} + i_{ac}\sin(W\tau) \quad (1)$$

Шунтирующая ёмкость, как и шунтирующий резистор подключены к джозефсоновскому переходу параллельно, тем самым изменяя значение сопротивления и ёмкости в уравнении (1) по соответствующим правилам. В правой части уравнения (1) суммарный ток, протекающий через джозефсоновский переход и, как следует из сказанного выше, шунтирующий конденсатор и резистор, состоит из двух

слагаемых. Второе слагаемое в данной записи представляет собой величину генерируемого переменного тока, частота которого фиксируется резонансной частотой резонатора. Таким образом, наличие резонатора в системе генератора обуславливается единственной составляющей переменного тока. Общая схема с учетом резонатора, показана на рис.1.

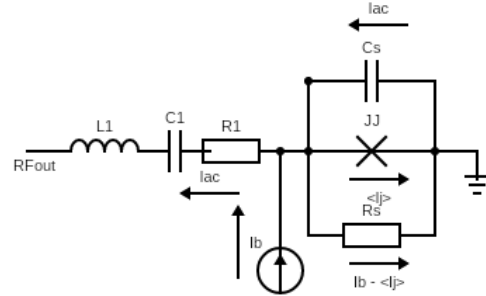


Рис.1. Общая схема генератора. R1, L1, C1 - параметры резонатора

В качестве критерия появления генерации переменного тока использовалось условие, описанное в работах [1], [2]. Для расчета импеданса использовался метод предложенный в [3], в нормированных величинах:

$$Z_w = R_w + jX_w = \left(\frac{1}{T i_w} \right) \int_0^{T \rightarrow \infty} \varphi e^{j\omega t} dt \quad (2)$$

В модели также было учтено, что получаемая мощность генерации расходуется на диссипацию и полезную нагрузку. Полную мощность потерь можно описать через сопротивление R1, которое связано с суммарной добротностью резонатора. Таким образом, было получено, что для появления устойчивой генерации переменного тока должно выполняться равенство мощностей, которое дополняет критерий появления устойчивой генерации, учитывая расход мощности на внешнюю нагрузку:

$$Re(Z_j(\omega)) = -R_1 \quad (3)$$

где $Re(Z_j(\omega))$ - действительное значение импеданса при падении переменного напряжения на джозефсоновском контакте.

Изготовление отдельных элементов генератора

Для новизны и упрощения изготовления генератора, используется планарная технология. В работе подразумевается использования джозефсоновского перехода, который представляет собой SNS контакт, так как изготовление такого контакта является менее затруднительным. Критический ток SNS контакта зависит от температуры, в связи с этим, были проведены измерения критического тока в зависимости от температуры с целью нахождения наиболее оптимального значения критического тока для работы генератора. Результат измерений показан на рис.2.

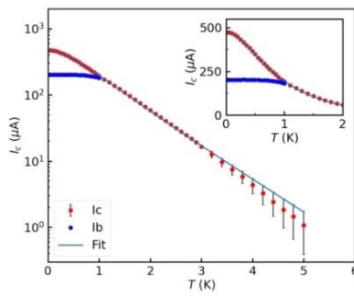


Рис.2. Зависимость критического тока от температуры для полученных SNS контактов

Экспериментально измерена ёмкость планарных конденсаторов, показанных на рис.3. Результат измерения полученной ёмкости составляет 40 пкФ, что соответствует малым значениям параметра маккамбера в уравнении (1).

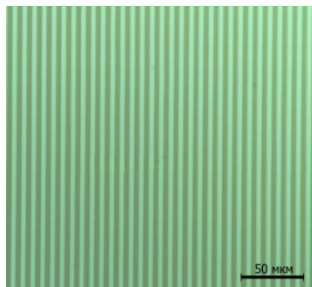


Рис.3. Увеличенное изображение одного из полученных планарных конденсаторов. Зеленым цветом сфотографирован металл, черным – подложка из кремния

Используя характерные значения критического тока для полученных SNS джозефсоновских переходов и значения ёмкостей для полученных планарных конденсаторов, было проведено моделирование получаемой мощности генератора. В качестве возможных переменных, от которых зависит выходная мощность, были выбраны критический ток и суммарное сопротивление джозефсоновского перехода с учетом шунтирующего резистора, так как эти параметры легче всего менять при изготовлении. Результат моделирования показан на рис.4. Максимальная мощность генератора оценивается в 15 пВт (показана красной точкой) и достигается при критическом токе 24 мкА и суммарном сопротивлении 2 Ома. Данной мощности хватает для управления кубитами.

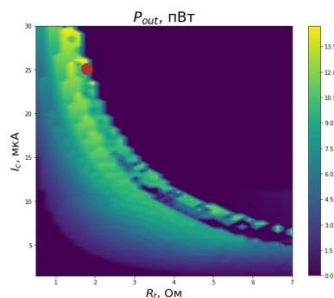


Рис.4. Зависимость мощность генерации переменного тока (обозначена цветом) от реальных параметров системы: I_c , R_t

Заключение

В результате работы приведен анализ возможности генерации при различных значениях параметров Маккамбера, соответствующей суммарной емкости системы, и частоты генератора, нормализованной к критической частоте джозефсоновского перехода, которая на эксперименте обуславливается частотой работы кубитов. В работе предложена более простая модель генератора, основанная на изготовлении чипа по планарной технологии, которая позволяет изготовить чип всего в две итерации. Были изготовлены и измерены образцы планарных конденсаторов (штырковых) и нормального сопротивления, представляющих собой полосу нормального металла. Используя измеренные характеристики, с помощью моделирования оценивается возможность генерации и мощность такого генератора.

Список литературы

1. Chengyu Y., Juha H., Visa V., Jinli Z., Joni I.Leif G., Jan G., Mikko M. A low-noise on-chip coherent microwave source. Nature Electronics. 2021. 885–892.
2. Hassel J., Grönberg L., Helistö P., Seppä H. Self-synchronization in distributed Josephson junction arrays studied using harmonic analysis and power balance. Appl. Phys. Lett. 2006. 89, 072503.
3. Zhai Z., Parimi P. V., Sridhar S. Nonlinear microwave impedance of short and long Josephson junctions. Physical Review B 59(14). 1999.

ДИССИПАЦИЯ И МНОЖЕСТВЕННЫЕ ХАРАКТЕРНЫЕ ДЛИНЫ ВО ВРЕМЯ ДИНАМИЧЕСКИХ СОБЫТИЙ В ГРЯЗНЫХ СВЕРХПРОВОДЯЩИХ ТОНКИХ ПЛЕНКАХ

Тейшейра Сарайва Т.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Введение

Сверхпроводящие мезоскопические системы обладают квантовым характером и в то же время могут быть в достаточной степени описаны с помощью эффективной теории, например теорией Гинзбурга-Ландау. В последнее десятилетие интерес к этим системам возрос в связи с появлением сверхбыстрых детекторов. В данной работе исследуются динамические конфигурации параметра порядка (удлиненные вихри) в сверхпроводящих тонких пленках с примесями при движения вихря. Образец - квадратная тонкая пленка с квадратным отверстием в центре, во внешнем перпендикулярном магнитном поле. Движение вихрей происходит, например, когда захваченный вихрь вырывается из нестабильного энергетического минимума в центре пленки при уменьшении приложенного магнитного поля.

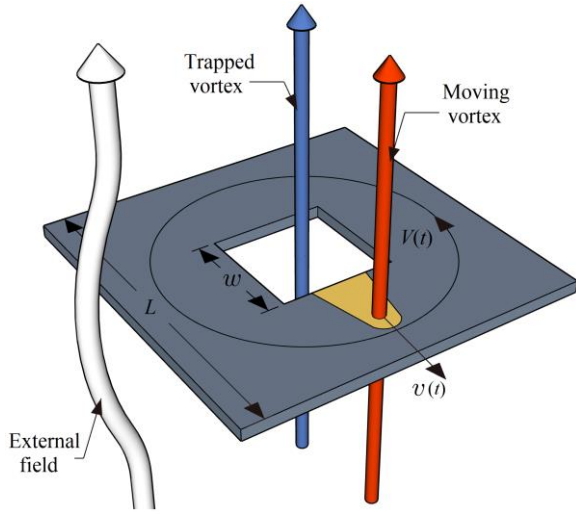


Рис.1. Набросок настройки исследуемой физической системы

Методы

Выполнялось компьютерное моделирование обобщенных нестационарных уравнений Гинзбурга-Ландау [1-3] в двумерной квадратной решетке для параметра порядка, векторного потенциала и локальной температуры:

$$\frac{u}{\sqrt{1+\gamma^2|\Psi|^2}} \left(\frac{\partial}{\partial t} + \frac{1}{2}\gamma^2 \frac{\partial |\Psi|^2}{\partial t} \right) \Psi = (\nabla - i\mathbf{A})^2 \Psi + (1 - |\Psi|^2) \Psi,$$

$$\frac{\partial \mathbf{A}}{\partial t} = \text{Im}[\Psi^*(\nabla - i\mathbf{A})\Psi] - \kappa_{\text{eff}}^2 \nabla \times \nabla \times \mathbf{A},$$

$$C_{\text{eff}} \frac{\partial T}{\partial t} = K_{\text{eff}} \nabla^2 T + \frac{1}{2} W_{\text{Total}} - \eta(T - T_0).$$

Рассеянное тепло выделяется за счет:

$$W_{\text{Total}} = 2 \left(\left| \frac{\partial \mathbf{A}}{\partial t} \right| \right)^2 + \frac{2u}{\sqrt{1+\gamma^2|\Psi|^2}} \left[\left| \frac{\partial \Psi}{\partial t} \right|^2 + \frac{\gamma^2}{4} \left(\frac{\partial |\Psi|^2}{\partial t} \right)^2 \right].$$

Наиболее важными вспомогательными параметрами системы являются: L - длина стороны пленки, а w - длина стороны отверстия. $\gamma = 3.6k_B\tau_{\text{e-ph}}/\hbar$ (а $\tau_{\text{e-ph}}$ - время электрон-фононного неупругого рассеяния), эффективный параметр Гинзбурга-Ландау $\kappa_{\text{eff}} = \lambda^2/d\xi$, (в единицах длины проникновения, длины когерентности и толщины пленки). C_{eff} и K_{eff} - это эффективные значения теплоемкости и теплопроводности пленки, а η - коэффициент теплопередачи подложки [3-4].

Результаты

Было проведено моделирование параметра порядка, векторного потенциала и локальной температуры во времени на основе уравнений GTDGL для пленки размером $L=26\xi(0)$ и $w=6\xi(0)$ и $w=4\xi(0)$ при следующих параметрах: $\gamma=0$, $T=0,91T_c$, $\kappa_{\text{eff}}=5$, $C_{\text{eff}}=0,65$, $K_{\text{eff}}=0,06$, $\eta=2 \times 10^{-4}$. Внешнее магнитное поле увеличивалось до тех пор, пока два вихря не проникли в образец и не застряли в центральном отверстии, затем поле уменьшали до тех пор, пока один из вихрей не покинул образец, оставив за собой след рассеянного тепла, и, следовательно, характер-

ная длина вихря увеличилась в направлении, противоположном движению вихря.

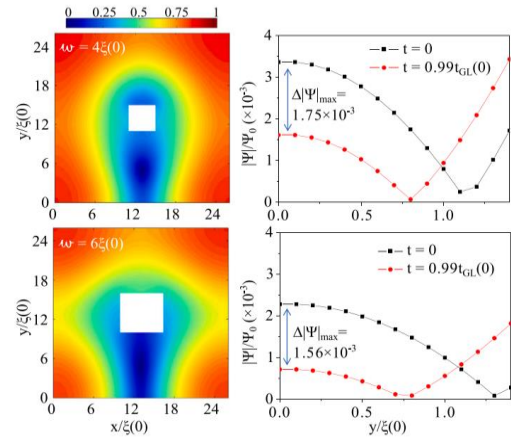


Рис.2. Слева - график $|\Psi|^2$ на пленке. Справа - $|\Psi|^2$ на средней линии пленки в два разных момента времени с разницей в $0,99t_{GL}(0)$

Заключение

В результате моделирования была получена динамика параметра порядка и магнитного поля внутри сверхпроводящей тонкой пленки во времени. Важно отметить, что используемые параметры достаточно далеки от точки Богомольного системы ($\kappa_{\text{eff}}=5$), и все же были обнаружены аномальные конфигурации параметра порядка. Многочисленные характерные длины, обнаруженные в системе, показывают увеличение нелокальности, и поэтому ожидается появление конфигураций, немного отличающихся от стандартных круговых вихрей. Это аналогично случаю с многозонными сверхпроводниками с различными скоростями Ферми [5].

Список литературы

1. Schmid, A. A., Time dependent Ginzburg-Landau equation and its application to the problem of resistivity in the mixed state. *Phys Kondens Materie* 5, 302–317 (1966) <https://doi.org/10.1007/BF02422669>
2. Kramer, L. and Watts-Tobin, R. J., Theory of Dissipative Current-Carrying States in Superconducting Filaments, *Phys. Rev. Lett.* 40, 1041 (1978) <https://doi.org/10.1103/PhysRevLett.40.1041>
3. G. R. Berdiyrov, M. V. Milošević, F. M. Peeters, Spatially dependent sensitivity of superconducting meanders as single-photon detectors, *Appl. Phys. Lett.* 100, 262603 (2012) <https://doi.org/10.1063/1.4731627>
4. E.C.S. Duarte, E. Sardella, T.T. Saraiva, A.S. Vasenko, R. Zadorosny, Comparing energy dissipation mechanisms within the vortex dynamics of gap and gapless nano-sized superconductors, *Materials Science & Engineering B*, 296, 116656 (2023) <https://doi.org/10.1016/j.mseb.2023.116656>
5. A. Vagov, A. A. Shanenko, M. V. Milošević, V. M. Axt, V. M. Vinokur, J. Albino Aguiar, and F. M. Peeters, Superconductivity between standard types: Multiband versus single-band materials, *Phys. Rev. B* 93, 174503 (2016) <https://doi.org/10.1103/PhysRevB.93.174503>

ИЗУЧЕНИЕ АТОМНЫХ СВОЙСТВ НА МЕЖЗЕРЕННОЙ ГРАНИЦЕ В ПЕРОВСКИТНЫХ МАТЕРИАЛАХ

Саматов М.Р.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В работе описывается миграция атомов на межзеренной границе в перовскитных материалах. Полученные результаты указывают на повышенную подвижность атомов брома (Br) в CsPbBr₃, что доказывает ранние предположения об ухудшении электрических свойств данных материалов из-за образования вакантных мест.

Введение

Металлогалогенидные перовскиты являются очень перспективными материалами для построения солнечных элементов нового поколения благодаря высоким оптоэлектронным свойствам и экономичностью их производства [1]. Эффективность преобразования энергии солнечных элементов на основе перовскитных материалов увеличилась с 3.8% в 2009 году до 26.1% в настоящее время, что приблизилось к рекордному показателю солнечных элементов на основе кремния [2]. Для дальнейшего увеличения производительности необходимо рассматривать влияние дефектов на электронные свойства, что и было показано в ряде экспериментальных работ [3-6]. Данные работы сосредоточены на ликвидации вредных структурных дефектов, служащих центрами рекомбинации носителей. Однако несмотря на малую предрасположенность к дефектам [7], в перовскитах наблюдается относительно быстрая рекомбинация носителей на границах зерен. На данный момент механизм влияния межзеренной границы на подвижность атомов, а также на электрические свойства и производительность солнечных элементов на основе перовскитов остается неясным [8]. Одной из ключевых идей данной работы является наличие атомов, перемещающихся вдоль межзеренной границы, как по некоторому коридору, что приводит к появлению вакантных мест.

Методы

В данной работе расчеты проводились с помощью систем машинного обучения на суперкомпьютере, расположенном в МИЭМ НИУ ВШЭ, а также высокоуровневого языка программирования Python для дальнейшей обработки результатов. Задание исходной структуры CsPbBr₃ производилось с помощью программы Atomsk [9], методом объединения двух пластин материала в противоположных направлениях. Полученная модель состоит из 6400 атомов и имеет размеры 47.46 x 53.05 x 106.14 Å. Расчет теории функциональной плотности (DFT) выполнялся с использованием пакета моделирования VASP [10]. Для описания электрон-ионных взаимодействий ис-

пользовался функционал Пердью-Берка-Эрнцера (PBE) [11] вместе с методом проектируемой расширенной полны (PAW) [12]. Анализ геометрической и электронной структуры модели производится методами машинного обучения, заранее обученного на пакете DeePMD-kit [13] и реализованного в коде LAMMPS для моделирования молекулярной динамики (MD). Все расчеты проводились при температуре 300 К и на временном промежутке 2нс (4000 временных точек). Дальнейшая постобработка и расчеты проводились с помощью языка программирования Python и встроенных в него библиотек.

Результаты

На рис.1 представлена схема рассматриваемой модели CsPbBr₃ в начальном состоянии, где красными точками представлены атомы Cs, синими – Pb и желтыми - Br. Все расчеты по обработке результатов проводились начиная с момента времени равного 0.1 нс, когда система переходит в стабильное состояние. В процессе стабилизации системы две части структуры сдвигаются в противоположных направлениях вдоль межзеренной границы. При этом в системе появляется продольное напряжение, которое сжимает структуру и может приводить к раннему разрушению материала.

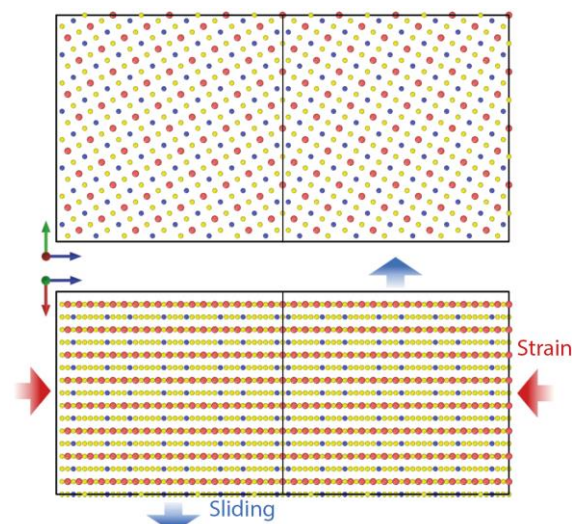


Рис.1. Схема рассматриваемой модели CsPbBr₃. Синими стрелками показаны направления сдвига, красными – направления напряжения

Результатами вычислений по симуляции структуры является массив координат атомов в каждый момент времени. Основной идеей данной работы было предположение, что повышенной подвижностью должны обладать атомы Br вдоль оси X (красная ось на рис.1), что должно быть видно при вычислении среднеквадратичного смещения (MSD). На рис.2 построена зависимость среднеквадратичного смещения атомов Br, рассчитанная для каждого момента времени, начиная с 0.1 нс. Постепенный рост графика с выходом на стабильные значения объясняется перемещением атомов внутри структуры с последующим занятием новых позиций. Как видно из данного гра-

фика предположение о преобладании движения вдоль оси X является неверным, следовательно мы можем рассматривать трехмерные перемещения для описания нашей структуры.

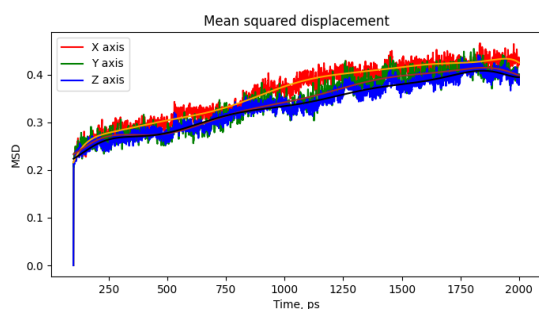


Рис.2. Среднеквадратичное смещение атомов Br как функция от времени. Красный график соответствует MSD для оси X, зеленый – оси Y и синий – оси Z

Однако среднеквадратичное отклонение не может показать нам отличие в подвижности атомов вдоль межзеренной границы и внутри цельного материала. Для этого была построена зависимость максимального смещения атомов брома, рассчитанная по всему временному промежутку (рис.3).

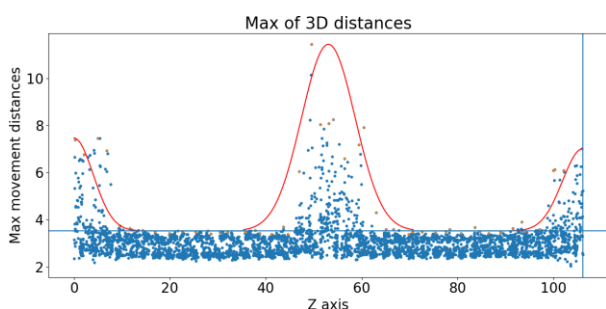


Рис.3. Максимальное перемещение атомов Br за весь временной промежуток

На рис.3 красной линией приведена огибающая для полученных результатов. В силу того, что наша структура является периодичной, на краях нашей системы также наблюдаются межзеренные границы. Из полученных результатов можно сделать вывод, что атомы Br имеют большую подвижность вблизи межзеренной границы и в пике достигает 7.2 \AA , что примерно соответствует 3 межатомным расстояниям в рассматриваемой структуре.

Для контроля эволюции системы необходимо рассматривать изменения в количестве подвижных атомов. Для этого введем понятие граничных атомов, как атомов, что имеют повышенную подвижность, т.е. за все время хотя бы раз перемещались на расстояние большее межатомного расстояния или являющихся более подвижными чем для структуры без дефектов. На рис.4 приведено среднеквадратичное смещение для атомов брома в структуре CsPbBr₃ без дефектов. В данном случае расчеты проводились для небольшой системы ($35.6 \times 35.6 \times 35.6 \text{ \AA}$), состоящей

из 1080 атомов и продолжительность симуляции составила 5.5 нс. Как мы можем видеть из полученных результатов чистая структура не нуждается в процессе стабилизации, т.е. отсутствуют атомы, перемещающиеся по структуре в течение времени.

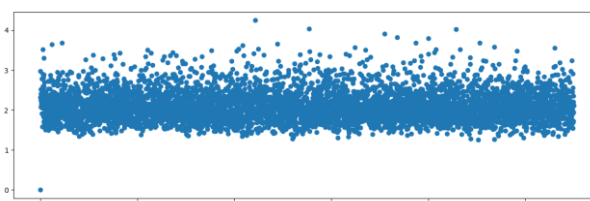


Рис.4. Среднеквадратичное смещение атомов Br как функция от времени для структуры без дефектов

Для наших вычислений максимальное значение (2.210408^2) на рис.4 является ключевым, т.к. оно отображает максимально возможное перемещение атомов брома в структуре без изменения своего положения (колебания относительно своего центра). Это значение было взято как минимальное для маркировки граничных атомов. На рис.5 представлена зависимость количества граничных атомов от времени. Логарифмический тип зависимости утверждает, что наша система со временем стабилизируется, т.е. есть начиная с некоторого момента времени изменения в структуре будут незначительными.

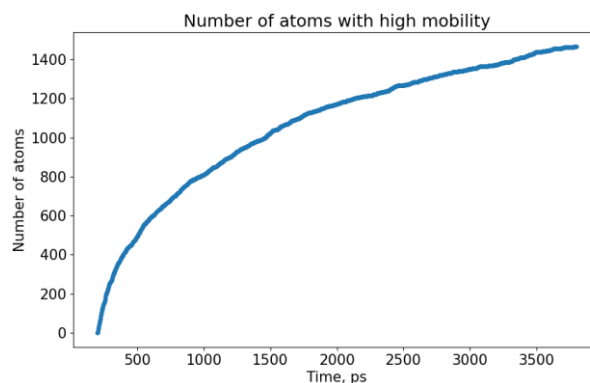


Рис.5. Зависимость количества граничных атомов (с высокой подвижностью) от времени

Заключение

В ходе данной работы было подтверждено наличие высокоподвижных атомов на межзеренной границе. Однако было опровергнуто предположение о большей подвижности атомов вдоль оси X (вдоль межзеренной границы). Рассмотрение максимального смещения атомов дает понимание о том на сколько далеко мигрируют высокоподвижные атомы. К примеру, для цельного идеального куска CsPbBr₃ максимальное значение смещения равно ~ 4 ангстрем, что соответствует размеру одной ячейки, а так как в идеальной структуре отсутствует миграция атомов мы можем рассматривать данное значение как граничное для высокоподвижных атомов. Рассмотрение смещения атомов Br в исходном образце

указывает на наличие атомов в структуре с максимальным отклонением большим 4 ангстрем, а значит самый подвижный атом «убегает» от своего начального положения на расстояние в 3 ячейки. Полученные результаты для эволюции системы указывают на предрасположенность системы к стабилизации. В прогнозе стабилизация произойдет при количестве высокоподвижных атомов равным 1600, что соответствует порядка половине всех атомов Вг в структуре. Однако данное предположение нуждается в тщательной проверке.

Список литературы

1. El-Ballouli A. O., Bakr O. M., Mohammed O. F. Compositional, Processing, and Interfacial Engineering of Nanocrystal- and Quantum-Dot-Based Perovskite Solar Cells // Chemistry of Materials - 2019. - Vol. 31, - P. 6387 - 6411.
2. Kojima A., Teshima K., Shirai Y., Miyasaka T. Organometal Halide Perovskites as Visible-Light Sensitizers for Photovoltaic Cells // Journal of the American Chemical Society - 2009. - Vol. 131, - P. 6050 - 6051.
3. Woo J. Y., Kim Y., Bae J., Kim T. G., Kim J. W., Lee D. C., Jeong S., Highly Stable Cesium Lead Halide Perovskite Nanocrystals through in Situ Lead Halide Inorganic Passivation // Chemistry of Materials - 2017. - Vol. 29, - P. 7088 - 7092.
4. Dong Q., Ho C. H. Y., Yu H., Salehi A., So F. Defect Passivation by Fullerene Derivative in Perovskite Solar Cells with Aluminum-Doped Zinc Oxide as Electron Transporting Layer // Chemistry of Materials - 2019. - Vol. 31, - P. 6833 - 6840.
5. Mannodi-Kanakkithodi A., Park J.-S., Jeon N., Cao D. H., Gosztola D. J., Martinson A. B. F., Chan M. K. Y., Comprehensive Computational Study of Partial Lead Substitution in Methylammonium Lead Bromide // Chemistry of Materials - 2019. - Vol. 31, - P. 3599 - 3612.
6. Maddala S., Chung C.-L., Wang S.-Y., Kollimalayan K., Hsu H.-L., Venkatakrishnan P., Chen C.-P., Chang Y. J. Forming a Metal-Free Oxidatively Coupled Agent, Bicarbazole, as a Defect Passivation for HTM and an Interfacial Layer in a p-i-n Perovskite Solar Cell Exhibits Nearly 20% Efficiency // Chemistry of Materials - 2020. - Vol. 32, - P. 127 - 138.
7. Kang J., Wang L.-W. High Defect Tolerance in Lead Halide Perovskite CsPbBr₃ // The Journal of Physical Chemistry Letters - 2017. - Vol. 8, -P. 489 - 493.
8. Lee J.-W., Bae S.-H., De Marco N., Hsieh Y.-T., Dai Z., Yang Y. The Role of Grain Boundaries in Perovskite Solar Cells // Materials Today Energy - 2018. - Vol. 7, - P. 149 - 160.
9. Hirel P. Atomsk: A Tool for Manipulating and Converting Atomic Data Files // Computer Physics Communications - 2015. - Vol. 197, - P. 212 - 219.
10. Kresse G., Hafner J. Ab Initio Molecular Dynamics for Liquid Metals // Physical Review B - 1993. - Vol. 47, - P. 558 - 561.
11. Perdew J. P., Burke K., Ernzerhof M. Generalized Gradient Approximation Made Simple // Physical Review Letters - 1996. - Vol. 77, - P. 3865 - 3868.

12. Blöchl P. E. Projector Augmented-Wave Method // Physical Review B - 1994. - Vol. 50, - P. 17953 - 17979.

13. Wang H., Zhang L., Han J., E W. DeePMD-Kit: A Deep Learning Package for ManyBody Potential Energy Representation and Molecular Dynamics // Computer Physics Communications - 2018. - Vol. 228, - P. 178 - 184.

КОМПЬЮТЕРНОЕ ИССЛЕДОВАНИЕ ПРИРОДЫ ФАЗОВЫХ ПЕРЕХОДОВ В СИСТЕМАХ ЗАРЯЖЕННЫХ ЧАСТИЦ, ОГРАНИЧЕННЫХ ПОТЕНЦИАЛОМ С КРУГОВОЙ СИММЕТРИЕЙ

Глуховцев П.И., Никонов Э.Г.
«Университет «Дубна»,

Институт системного анализа и управления
кафедра распределенных
информационно-вычислительных систем

Аннотация

В данной статье с помощью метода молекулярной динамики исследуются фазовые переходы (ФП) в системах заряженных частиц, расположенных на плоскости и ограниченных потенциалом с круговой симметрией. Авторы анализируют изменения фазового состояния в зависимости от числа частиц в условиях экстремально низких температур, приближающихся к абсолютному нулю.

Введение

Изучается система из N идентичных заряженных частиц с кулоновским взаимодействием в двумерном ограничивающем потенциале радиуса R . Гамильтониан такой системы записывается как:

$$H = \sum_{i=1}^N V(r_i) + \alpha \sum_{\substack{i,j=1 \\ i < j}}^N \frac{1}{|\vec{r}_i - \vec{r}_j|} + \sum_{i=1}^N T_i,$$

где $r_i = |\vec{r}_i|$ – это расстояние до центра области, ограниченной потенциалом, $\alpha = e^2/4\pi\epsilon_0\epsilon_r$ – величина, характеризующая силу взаимодействия зарядов в среде, T_i – кинетическая энергия частицы. Ограничивающий потенциал $V(r)$ определяется как:

$$V(r) = \begin{cases} 0, & r < R \\ \infty, & r \geq R. \end{cases}$$

Подобного рода системы с различным потенциалом встречаются в таких природных системах как: вихри Абрикосова, вихревая решетка во вращающемся конденсате Бозе-Эйнштейна или кластер вихрей электронной плотности в квантовой точке. Эти явления могут использоваться, например, при создании полупроводников, сверхпроводников или лазеров, а исследование ФП направлено на определение критических свойств веществ.

В рамках исследования рассматривается переход Березинского-Костерлица-Таулеса (БКТ) — это пример топологического фазового перехода, который обнаруживается в двухмерных системах. Это уникальный тип фазового перехода, который не может быть описан обычной теорией Ландау фазовых переходов. Вместо того, чтобы происходить при определенной температуре, как в теории Ландау, переход БКТ связан с возникновением топологических дефектов в системе [1].

Для обнаружения топологических фазовых переходов такого типа можно использовать конфигурационные характеристики, которые обозначают нарушение упорядоченности в двумерных системах, такие как ориентационный параметр порядка, топологический заряд, ориентационная и трансляционная корреляционные функции.

Ориентационный параметр порядка

Для описания симметрии исследуемой системы используется параметр порядка, описывающий круговую симметрию 6 порядка Ψ_6 :

$$\Psi_6(\mathbf{r}_i) = \frac{1}{n(i)} \sum_{j=1}^{n(i)} e^{i6\theta_{ij}},$$

где θ_{ij} — угол между вектором \mathbf{r}_{ij} , соединяющим i -ю и j -ю частицы, и произвольной осью (см. рис.1 жирная линия), $n(i)$ — число соседей i -й частицы. Таким образом каждой частице в соответствие ставится некоторый вектор.

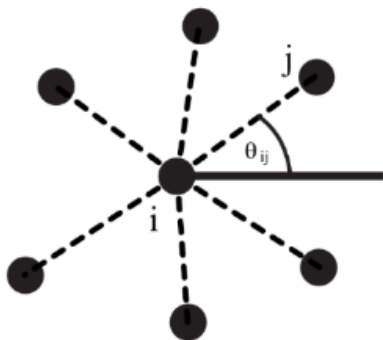


Рис.1. j ближайших соседей i -й частицы

Число соседей $n(i)$ определяется по построению Вороного.

В результате для каждой частицы при определении длины вектора получаем значение в диапазоне $[0,1]$. Ближкое к единице число в данном случае будет означать высокую гексагональную симметрию, а меньшее значение ее нарушение. При установлении некоторого порога разделения можно классифицировать принадлежность частицы к узлам гексагональной решетки (см. рис.2).

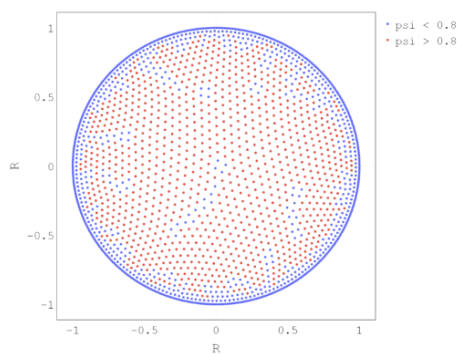


Рис.2. Ориентационный параметр порядка системы из $N=2000$ частиц. Красным цветом отсечены частицы со значением $|\Psi_6| > 0.8$

Сам по себе данный ориентационный параметр порядка не служит сигналом фазового перехода системы и требуется для вычисления других функций. Однако распределение этой величины может нести полезную информацию о нарушении гексагональной решетки системы в общем [2]. Например, для системы с меньшей энергией и наличием большего количества упорядоченных частиц распределение $|\Psi_6|$ смещается к 1 (см. рис.3,4).

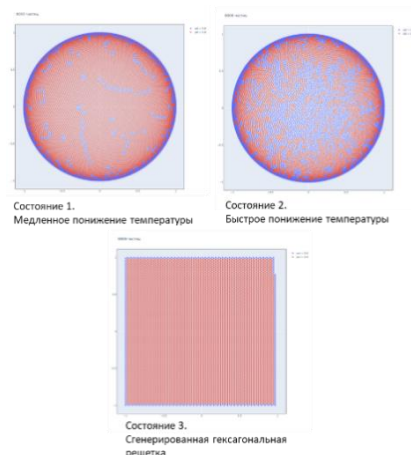


Рис.3. Пример систем в различных состояниях для $N=8000$. Состояние 1 — минимально найденная энергия, состояние 2 — высокая потенциальная энергия, состояние 3 — сгенерированная гексагональная решетка

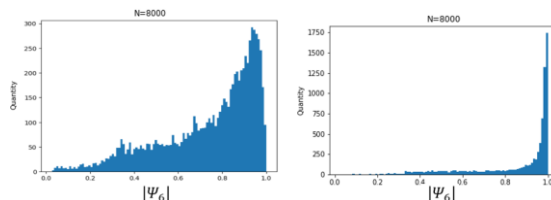


Рис.4. Распределение $|\Psi_6|$ для состояния 1 (справа) и состояния 2 (слева)

Ориентационная корреляционная функция

Для определения сохранения упорядоченности гексагональной фазы системы на определенном со-

стоянии используется ориентационная корреляционная функция G_6 :

$$G_6(r) = \left\langle \sum_{i \neq j} \Psi_6^i \Psi_6^{j*} \delta(r - |\mathbf{r}_i - \mathbf{r}_j|) \right\rangle,$$

где $|\mathbf{r}_i - \mathbf{r}_j|$ расстояние между i -ой и j -ой частицами, δ — дельта функция [4]. Таким образом получаем среднее алгебраическое значение Ψ_6 для всех пар частиц, расстояние между которыми меньше либо равно r . Для системы с кристаллической фазой график ведет себя как константа. При гексатической фазе типичное поведение — это затухание степенному алгебраическому закону. В изотропной фазе корреляции затухают экспоненциально быстро. Для исследуемой замкнутой системы данный параметр ограничен диаметром области моделирования (диаметр равен $2R$), т.к. после превышении параметра r этой величины G_6 принимает нулевое значение. Величина r представлена в единицах радиуса R . Трудности возникают и при определении одновременного существования двух фаз на границе и в центре системы (см. рис.5).

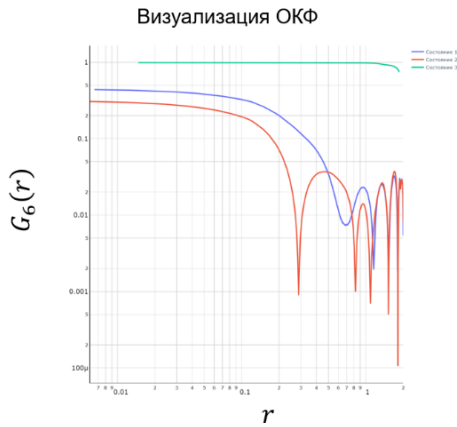


Рис.5. Ориентационная корреляционная функция G_6 для состояний системы с разной упорядоченностью убывает по разным законам

Другой вариант ориентационной корреляционной функции может характеризовать сохранение упорядоченности внутри окрестности вокруг частицы:

$$G_6^*(r) = \left\langle \sum_{i \neq j} \Psi_6^i \Psi_6^{j*} \theta(r - |\mathbf{r}_i - \mathbf{r}_j|) \right\rangle,$$

где θ — ступенчатая функция Хевисайда. В данном случае при превышении параметром r диаметра области моделирования функция перестанет изменяться и примет значение среднего $|\Psi_6|$ по всем частицам. Оценивать эту функцию можно как изменение средней симметрии подобласти в зависимости от ее размера.

Топологический заряд

Следующий вариант возможного сигнала фазового перехода требует некоторого определения. Будем

считать, что топологический заряд частицы равен $(b-r)$, где r — число ближайших соседей (определенного по построению Вороного) (см. рис.6) [3,4]. В таком случае можно рассмотреть распределение зарядов и суммарный топологический заряд системы по всем частицам. Для построения распределения используем все экспериментальные данные с различными начальными и конечными состояниями, группируя их по числу частиц.

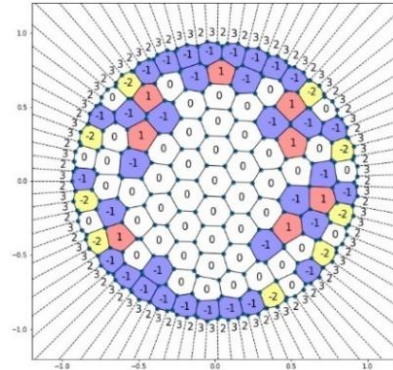


Рис.6. Пример топологического заряда частиц в областях разбиения Вороного для $N=187$

Следующим шагом необходимо определить некоторое фазовое пространство признаков и разбить его на подобласти, соответствующие искомым физическим фазам. Например, для этого было рассмотрена зависимость суммарного топологического заряда от энергии основного состояния для различных начальных условий (см. рис.7).

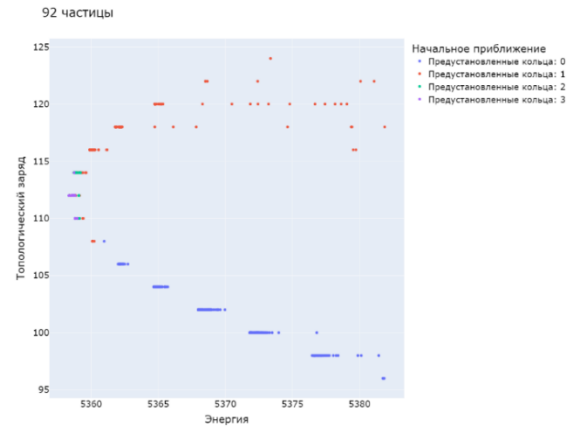


Рис.7. Статистическая зависимость суммарного топологического заряда от энергии основного состояния для $N=92$

При исследовании было обнаружено что с ростом числа частиц в системе возникают состояния с одинаковым суммарным топологическим зарядом, но значительно различающейся энергией. Использование приведенной диаграммы для определения фазы вещества требует дальнейшего определения фазы в каждой точке и большего количества статистических данных.

Проведение расчетов

Для численного интегрирования использовался алгоритм Бимана с порядком точности 3. Метод в явной форме записывается следующим образом:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{4\mathbf{a}(t) - \mathbf{a}(t - \Delta t)}{6}\Delta t^2,$$
$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{5\mathbf{a}(t) + 2\mathbf{a}(t + \Delta t) - \mathbf{a}(t - \Delta t)}{6}\Delta t,$$

где $\mathbf{r}(t)$ – радиус-вектор, $\mathbf{v}(t)$ – скорость, $\mathbf{a}(t)$ – ускорение частицы в момент времени t .

Расчеты проводились на языке программирования *python* с использованием библиотеки *torch*. Ускорение достигалось за счет параллельных векторизованных вычислений сил и интегрирования на GPU (см. рис.8).

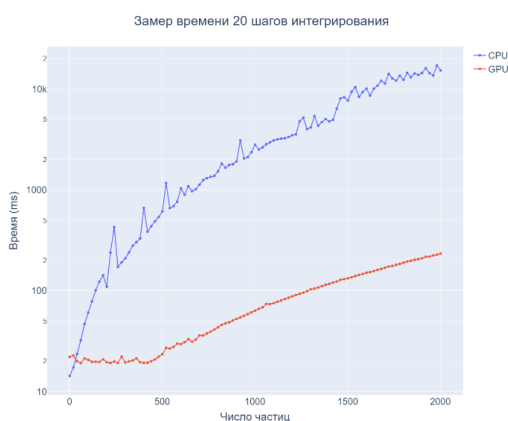


Рис.8. График сравнительного времени выполнения 20 шагов интегрирования на GPU (Tesla T4 16GB) и CPU (Intel® Xeon® CPU @ 2.00GHz)

Заключение

В рамках проведенного исследования был выполнен анализ эффективности ориентационного параметра порядка и соответствующей корреляционной функции для поиска сигналов фазовых переходов в двумерных системах заряженных частиц, ограниченных круговым потенциалом, при увеличении числа частиц. Это, в свою очередь, может привести к более глубокому пониманию упомянутых выше систем и реакции их на фазовые изменения. Применение молекулярно-динамического моделирования для анализа стационарных состояний систем при различном количестве частиц обеспечило необходимое количество данных для исследования. Результаты работы могут быть использованы для улучшения экспериментальных методов исследования и компьютерного моделирования физических процессов, а также расширения теоретической базы о фазовых переходах в физических системах.

Список литературы

1. Рыжов В. Н., Тареева Е. Е., Фомин Ю. Д., Циок Е. Н. // УФН, 2017, Т. 187, № 9, С. 921.
2. Jami, P., Ash, V. & Ghosal, A. Significance of the nature of disorder on the universal features of the spatio-

temporal correlations of two-dimensional Coulomb-clusters. Eur. Phys. J. B 96, 45 (2023). <https://doi.org/10.1140/epjb/s10051-023-00510-5>.

3. Qingyou Meng and Gregory M. Grason // Physical Review E104, 034614 (2021)

4. Kong M., Partoens B., Peeters F. M. // Phys. Rev E, 2003, V.67, P. 021608.

ИССЛЕДОВАНИЕ ВЛИЯНИЯ СТРУКТУРНОГО БЕСПОРЯДКА НА ШИРИНУ ЗАПРЕЩЕННОЙ ЗОНЫ ГАЛОИДНОГО ПЕРОВСКИТА Cs₂AgBiBr₆

Абзах Б.А., Васенко А.С.

Национальный исследовательский университет
«Высшая школа экономики»
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В работе рассматривается влияние разупорядочения структуры галоидного перовскита Cs₂AgBiBr₆ на ширину его запрещенной зоны. Данное исследование проводится в рамках разработки перспективных солнечных элементов на основе перовскитов для использования в зеленой энергетике.

Введение

Применение двойных галогенидных перовскитов как материалов для использования в солнечных элементах имеет на сегодняшний день многообещающие перспективы. Конкретно, перовскит Cs₂AgBiBr₆ на сегодняшний день представляется одним из наиболее удачных кандидатов на роль основного материала для солнечных элементов, благодаря своей химической стабильности и нетоксичности – качествам, которые выгодно отличают его от перовскитов, имеющих в своем составе свинец [1]. Однако, главной проблемой при внедрении этого материала в солнечную энергетику является его низкая энергетическая эффективность, которая обусловлена, в первую очередь, слишком большой шириной его запрещенной зоны (около 2 эВ). Для решения этой проблемы требуется понимание физических механизмов, влияющих на ширину запрещенной зоны.

Так недавно проведенные эксперименты показали резкое уменьшение ширины запрещенной зоны в Cs₂AgBiBr₆, подвергнутом умеренному (без плавления образца) нагреву. Это явление связывают с перераспределением атомов Ag и Bi в кристаллической решетке материала [2, 3]. Тем не менее, четкое понимание как связаны между собой неупорядоченность структуры этого материала и уменьшение ширины его запрещенной зоны до сих пор отсутствует.

Целью данной работы являлось определение механизма влияния структурного беспорядка на ширину запрещенной зоны перовскита Cs₂AgBiBr₆. При этом мы исходили из предположения, что основным механизмом, который обеспечивает уменьшение ши-

рины запрещенной зоны перовскита при хаотизации его структуры является неупорядоченное распределение в его решетке атомов Ag и Bi. Этот механизм реализуется посредством орбитальной гибридизации электронов этих атомов.

Использованные методы расчета электронной структуры

Для расчета свойств перовскита $\text{Cs}_2\text{AgBiBr}_6$ с неупорядоченной структурой нами были проведены расчеты методом функционала плотности (DFT) с использованием программного пакета VASP. Расчеты проводились для элементарной ячейки, содержащей 40 атомов. Разупорядочение структуры материала создавалось сменой положения атомов Ag и Bi внутри элементарной ячейки. Для описания электронного взаимодействия использовался функционал PBE (Perdew-Burke-Ernzerhof) и потенциалы типа PAW (projector-augmented wave). Из-за того, что при использовании функционала PBE обычно существенно занижается значение ширины запрещенной зоны полупроводника, мы так же использовали в расчетах гибридный функционал HSE+SOC (Heyd-Scuseria-Ernzerhof + spin-orbital coupling) с учетом спин-орбитального взаимодействия [4].

Результаты расчетов

На рис.1 показана плотность электронных состояний (DOS – density of states) перовскита $\text{Cs}_2\text{AgBiBr}_6$ с упорядоченной структурой, рассчитанная с использованием двух функционалов: 1) PBE и 2) HSE+SOC. Ширина запрещенной зоны, рассчитанная при помощи функционала типа PBE, составила 1.33 эВ, что существенно меньше значения, экспериментально измеренного для этого материала (около 2 эВ). Использование функционала типа HSE+SOC дает существенно более близкое к эксперименту значение – 1.96 эВ, что подтверждает надежность использования этого метода для расчетов электронной структуры этого материала. Парциальная плотность состояний (плотность состояний, отнесенная к каждому элементу) показывает, что максимум валентной зоны (VBM – valence band maximum) определяется в основном вкладами атомов Ag и Br, а минимум зоны проводимости (CBM – conduction band minimum) – вкладом атомов Bi.

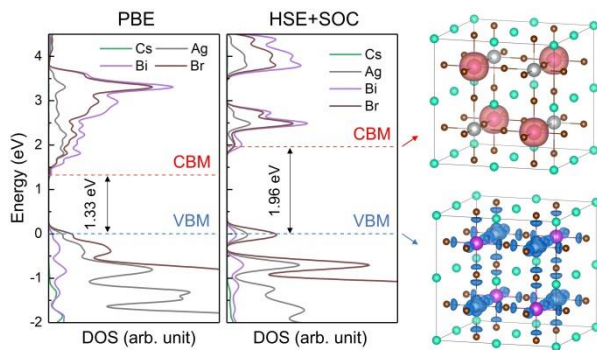


Рис.1. Плотность состояний (DOS) и плотность заряда для $\text{Cs}_2\text{AgBiBr}_6$ с упорядоченной структурой

Два типа разупорядочения в $\text{Cs}_2\text{AgBiBr}_6$ конструировались путем изменения положения атомов Ag и Bi в элементарной ячейке. Рассчитанная плотность состояний для этих случаев показана на рис.2. Как видно из рисунка, ширина запрещенной зоны существенно снизилась (вплоть до 0.08 эВ) для разупорядоченных структур, что подтверждает сделанное предположение. Плотность распределения заряда, так же показанная на этом рисунке, показывает изменение орбитальной гибридизации.

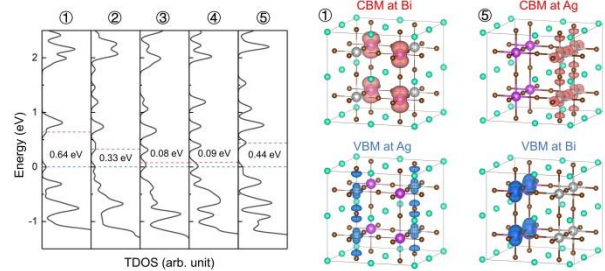


Рис.2. Плотность состояний и плотность заряда для $\text{Cs}_2\text{AgBiBr}_6$ с неупорядоченной структурой

Стоит заметить, что рассчитанные нами величины для ширины запрещенной зоны разупорядоченного $\text{Cs}_2\text{AgBiBr}_6$ сильно меньше экспериментально наблюдаемых, которые лежат в районе 1.7 эВ [2].

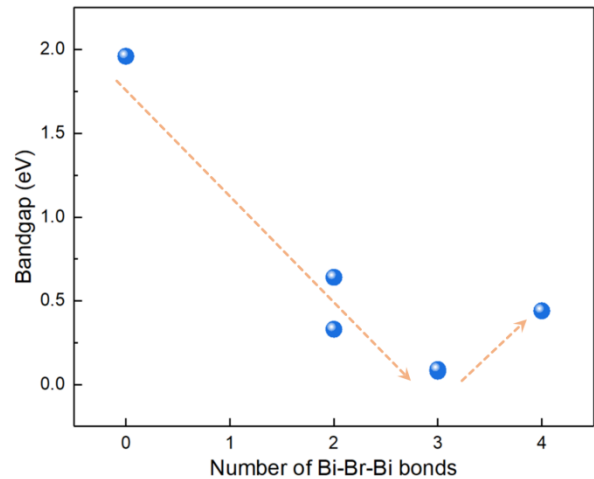


Рис.3. Корреляция между количеством связей и шириной запрещенной зоны

На этом графике мы грубо наблюдаем отрицательную корреляцию между количеством связей Bi-Br-Bi в неупорядоченных структурах и результирующими запрещенными зонами, однако эта корреляция не является монотонной, поскольку в модели с 4 связями мы наблюдаем большую запрещенную зону, чем в предыдущей структуре (0.44 эВ), из-за к обмену состояниями VBM и CBM на рис.3.

По нашему мнению, это расхождение может быть вызвано недостаточным объемом элементарной ячейки, использованной в расчетах, что вызвало слишком большую степень беспорядка. В дальнейшем мы планируем использовать большую по размерам элементарную ячейку для более реалистичного

представления разупорядоченной структуры $\text{Cs}_2\text{AgBiBr}_6$, что, впрочем, приведет к увеличению временных затрат на проведение расчетов.

Заключение

В нашей работе расчетным путем получен качественный результат, состоящий в том, что разупорядочение кристаллической структуры перовскита $\text{Cs}_2\text{AgBiBr}_6$ приводит к уменьшению ширины его запрещенной зоны, причем основной вклад в этот процесс вносит изменение положения атомов Ag и Bi в элементарной ячейке. Для получения более реалистичных количественных результатов необходимо увеличить кристаллическую ячейку, используемую в расчетах.

Список литературы

1. Tress W., Sirtl M.T. $\text{Cs}_2\text{AgBiBr}_6$ double perovskites as lead-free alternatives for perovskite solar cells. // Sol. RRL. -2022. -№6, -p. 2100770.
2. Ji F. et al. Lead-free halide double perovskite $\text{Cs}_2\text{AgBiBr}_6$ with decreased band gap. // Angew. Chem. Int. Ed. -2020. № 59, -p. 15191.
3. Zhang Y. et al. Thermochromic $\text{Cs}_2\text{AgBiBr}_6$ single crystal with decreased band gap through order-disorder transition. // Small. -2022. № 18, -p. 2201943.
4. Xiao Z., Meng W., Wang J., Yan Y. Thermodynamic stability and defect chemistry of bismuth-based lead-free double perovskites. // ChemSusChem. -2016. -№9, -p. 2628.

СВЕРХПРОВОДЯЩИЙ ДИОДНЫЙ ЭФФЕКТ В ТОПОЛОГИЧЕСКИХ ГИБРИДНЫХ СТРУКТУРАХ

Карабасов Т., Васенко А.С.

*Национальный исследовательский университет «Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе исследован сверхпроводящий диодный эффект в Джозефсоновском S/TI/S (сверхпроводник/топологический изолятор/сверхпроводник) переходе. В работе были продемонстрированы зависимости эффективности диода от параметров системы.

Введение

Сверхпроводящий диодный эффект является активной областью исследования в физике конденсированного состояния. В частности, трехмерные топологические изоляторы (ТИ) могут оказаться одним из наиболее подходящих материалов для реализации сверхпроводящих диодов. В низкоэнергетическом разложении Гамильтониана ТИ принято учитывать только линейные и квадратичные члены. Обычно эффектом гексагональных искажений решетки пренебрегают. Однако, роль гексагональных искажений может быть очень существенной при учете транс-

портных свойств топологических изоляторов, таких как Bi_2Se_3 или Bi_2Te_3 . В данной теоретической работе мы представляем исследование диодного эффекта Джозефсона на основе слабой связи топологического изолятора. Мы обращаемся к вопросу о влиянии гексагональной деформации на эффект диода Джозефсона. Мы утверждаем, что варпинг приводит к анизотропии эффекта диода Джозефсона.

Методология и модель

Модель, лежащая в основе построения сверхпроводящего диода, строится с помощью эффективного Гамильтониана, который можно записать в следующем виде:

$$H(k) = H_0 + H_{SOC} + H_w + H_s + H_z = \left(\frac{\hbar^2}{2m_{eff}}(k_y^2 + k_x^2) - \mu_x \right) \hat{s}_0 \otimes \hat{\sigma}_z + \alpha(k_y \hat{s}_x - k_x \hat{s}_y) \otimes \hat{\sigma}_z + \lambda k_x(k_x^2 - 3k_y^2) \hat{s}_z \otimes \hat{\sigma}_z + B \cdot \hat{s} \otimes \hat{\sigma}_z + \Delta \hat{s}_0 \otimes \hat{\sigma}_x. \quad (1)$$

Здесь μ_x — химический потенциал в соответствующей области ($x = S$ для сверхпроводящей области S и $x = N$ для поверхности топологического изолятора TI), α — сила спин-орбитального взаимодействия, а λ — коэффициент варпинга. $B = (B_x, B_y, 0)$ — зеемановское поле. Параметр порядка Δ является комплексной функцией и может зависеть от координат. Матрицы \hat{s} и $\hat{\sigma}$ являются матрицами Паули в спиновом и частично-дырочном пространствах соответственно. Эти пространства объединяются кронекеровским произведением \otimes .

В данной работе будет рассмотрен случай S/TI/S структуры, которая показана на рис.1.

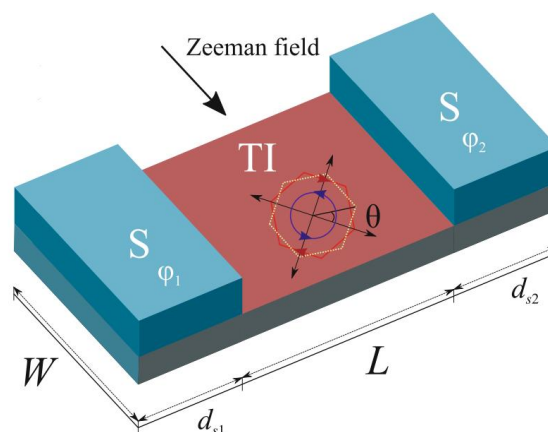


Рис.1. Схематичное изображение Гибридной структуры

Одна из важнейших характеристик диодного эффекта.

– эффективность диода. В этой работе используется следующее определение:

$$\eta = \frac{I_c^+ - |I_c^-|}{I_c^+ + |I_c^-|}, \quad (2)$$

где $I_c^+ = \max[I(\phi)]$ — критический ток Джозефсона в положительном направлении, а $I_c^- = \min[I(\phi)]$ — критический ток Джозефсона в противоположном направлении.

В следующем разделе представлены результаты расчета эффективности диода для заданной структуры.

Результаты

Для получения зависимости использовалась программа, написанная на языке программирования Python с использованием библиотеки Kwant [1].

На рис.2 приведен результат расчета диодной эффективности в зависимости от поля B_y . Кривые рассчитаны при двух значениях сверхпроводящей щели Δ , что соответствует случаю короткого и длинного контакта. Из рисунка можно видеть, что в случае короткого контакта эффективность диода принимает отрицательные значения, при этом перемены знака не происходит. Однако, в случае длинного Джозефсоновского контакта эффективность может изменить знак при определенном значении поля.

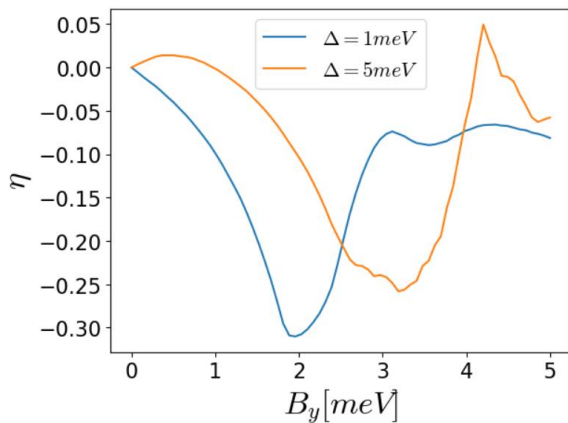


Рис.2. График эффективности сверхпроводящего диода от магнитного поля в плоскости. На рисунке представлены две кривые: при $\Delta = 1$ meV (синяя линия) и при $\Delta = 5$ meV (оранжевая линия)

Заключение

В ходе работы был изучен диодный эффект в Джозефсоновском контакте, где в качестве слабой связи использована поверхность топологического изолятора. В рамках формализма уравнений Боголюбова – Де Жена была построена модель гибридной структуры с учетом гексагональных искажений решетки топологического изолятора. Была рассчитана диодная эффективность в зависимости от магнитного поля. В результате было показано качественно разное поведение эффективности для короткого и длинных контактов.

Список литературы

1. Groth C. W., Wimmer M., Akhmerov A. R., Waintal X. Kwant: a software package for quantum transport [Текст] / Groth C. W., Wimmer M., Akhmerov A. R., Waintal X. //New Journal of Physics. — 2014. — № 16. — С. 063065.

Секция «Кибербезопасность»

ПРИНЦИП РАБОТЫ ДЕРЕВЬЕВ ВЕРКЛА И ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ИХ ПРИМЕНЕНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧ ЗАЩИТЫ ИНФОРМАЦИИ

Сербаев В.Р.

*Национальный исследовательский университет
«Высшая школа экономики»,
аспирантская школа по техническим наукам
МИЭМ НИУ ВШЭ*

Аннотация

В работе рассматривается принцип построения и работы относительно новой криптографической структуры дерева Веркла. Приводится описание имеющихся на сегодняшний день вариаций деревьев Веркла. Проводится сравнение с деревьями Меркла: выделены отличия и преимущества модели дерева Веркла перед моделью дерева Меркла. Дается обзор существующих решений в части применения деревьев Веркла. Оцениваются перспективы использования данных структур для решения задач в области защиты информации.

Введение

В 2018 году информационная экономика была внесена в перечень основных направлений стратегического развития Российской Федерации до 2025 года.

В условиях стремления нашей страны к устойчивому развитию, обеспечению технологического суверенитета на фоне экономических санкций с последовавшими за ними финансовыми и банковскими ограничениями возникает потребность в применении автономных инструментов цифровой экономики, среди которых называются системы распределенного реестра.

На сегодняшний день технология распределенного реестра может стать ключевым элементом для обеспечения автономности финансовых транзакций с учетом её успешного внедрения и применения в крупных государственных и частных компаниях различных секторов экономики [1]. Так, Цифровой Рубль, выпускаемый Банком России, работает на основе блокчейн-технологии, являющейся частным случаем системы распределенных реестров [2].

С развитием информационных технологий увеличилась не только потребность в безопасном и эффективном обмене данными, но и в возможности реализации таких обменов.

В блокчейн-системах для быстрой и эффективной верификации транзакций применяются деревья Меркла.

Принцип работы деревьев Меркла

Дерево Меркла, также известное как хеш-дерево – это полное двоичное дерево, в листовые вершины которого помещены хеши от блоков данных. Внут-

ренние вершины содержат хеши от сложения значений в дочерних вершинах. Корневой узел дерева содержит хеш от всего набора данных.

Хеш-деревья позволяют строить доказательство Меркла для обеспечения целостности данных. В одноранговых сетях такое доказательство гарантирует, что блоки данных, полученные от других одноранговых узлов, не были изменены, либо повреждены. Хеш-деревья также позволяют проверить, что другие одноранговые узлы не отправляют поддельные блоки.

Блокчейн состоит из цепочек блоков. Каждый блок может содержать до N транзакций. В качестве блоков данных для дерева Меркла в блокчейн-системе выступают транзакции. Корневой хэш, полученный в конце дерева Меркла, суммирует все транзакции, содержащиеся в данном блокчейн-блоке.

Для верификации транзакции требуется доказательство Меркла, включающее корень хеш-дерева и ветку, включающую хеши от верифицируемой транзакции до корня, позволяя не использовать весь объем информации о транзакциях, обращаясь к конкретной транзакции.

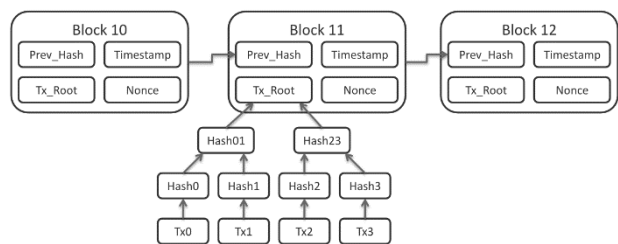


Рис. 1. Схема построение корня Меркла
в блок-чейн-сети

Так, доказательством включения транзакции в блок для транзакции Tx_2 будет $(Hash3, Hash01)$. Выполняется следующее вычисление корневого хеша:

$$\text{hash}(\text{Hash01}, \text{hash}(\text{Hash2}, \text{Hash3})) = \text{RootHash}$$

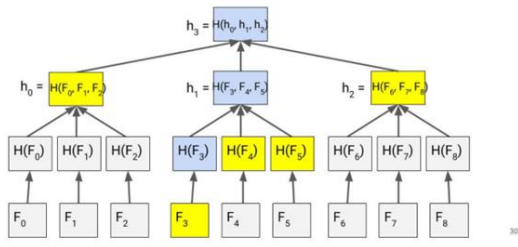
Далее RootHash сравнивается с Tx_Root . Данный подход делает невозможным подделку доказательства.

В общем случае сложность алгоритма доказательства Меркла оценивается как $O(K) = \log_2(N)$ действий, где K – длина пути Меркла (или же высота дерева), N – количество транзакций.

Применения структуры дерева Меркла имеет существенный недостаток: дерево Меркла, содержащее значительное число блоков, может иметь доказательство, которое будет непомерно большим. Так, при наличии 2^{30} ($\sim 10^9$) транзакций длина пути возрастет до 30. При использовании функции хэширования SHA-256 доказательство Меркла уже будет весить примерно один килобайт. Для большо-

го числа транзакций доказательство Меркла может создать дорогостоящие накладные расходы на пропускную способность. Поэтому с ростом блоков в дереве Меркла возникает проблема его эффективно-го обхода.

Деревья Меркла могут иметь коэффициент ветвления $l > 2$ (каждый узел имеет $l > 2$ дочерних узлов). В бинарном дереве Меркла доказательство состоит из одного узла на каждом уровне.



Доказательство выделено желтым цветом

Рис.2. Доказательство в l -ичном дереве хеширования для листа со значением F_3

Интуитивно может показаться, что, для того чтобы уменьшить размер доказательства, нужно уменьшить высоту дерева, придав ему коэффициент ветвления $l > 2$.

К сожалению, в таком случае доказательство Меркла наоборот становится только больше, увеличиваясь с $O(\log_2 n)$ до $O(l \log_2 n)$ (где n – количество блоков дерева). Это связано с тем, что в l -ичном дереве Меркла доказательство Меркла фактически состоит из $(l - 1)$ узлов на каждом уровне за исключением первого.

Принцип работы деревьев Веркла

Деревья Веркла выполняют функцию, аналогичную функции хеш-деревьев Меркла: всё так же можно разместить и хранить большое количество блоков в древовидной структуре и проводить вычисления короткого доказательства целостности / принадлежности любого отдельного блока или группы блоков, зная корень дерева.

Концепция дерева Веркла заключается в совмещении двух криптографических структур – дерева Меркла и схемы векторных обязательств (от англ. «Vector Commitment» + «Merkle» = «Verkle»). Основная идея дерева Веркла – построение l -ичного дерева Меркла, в котором значения узлов формируются посредством использования криптографических обязательств.

Построение оригинального дерева Веркла, использующего векторные обязательства, начинается с выбора коэффициента ветвления l . Имеющиеся блоки дерева $F_0, F_1, F_2 \dots, F_n$ распределяются на группы-подмножества по l блоков в каждой. Для каждого из подмножеств вычисляется векторное обязательство C_j . Для каждого обязательства вычисляются доказательства членства элементов π_i постоянного размера каждого подмножества для каждого F_i блока. Процесс вычисления криптографических обя-

зательств происходит снизу вверх по дереву по аналогии с вычисленными l обязательствами, но по l обязательствам вместо исходных блоков данных до тех пор, пока не будет вычислено корневое обязательство C_{root} .

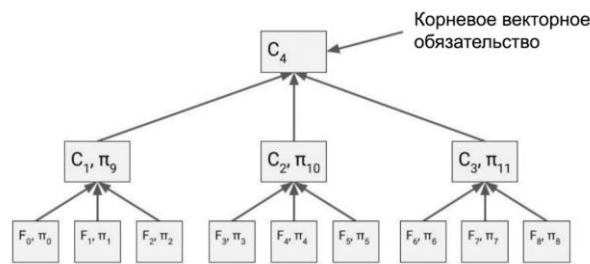


Рис.3. Схематичное изображение построения дерева Веркла с коэффициентом ветвления $l = 3$

Доказательство членства в дереве Веркла для выбранного блока-транзакции осуществляется посредством проверки доказательств членства π_i , проходя путь от блока, целостность которого необходимо установить, до корневого обязательства.

Так, доказательством включения блока в дерево Веркла для блока F_2 будет выглядеть следующим образом:

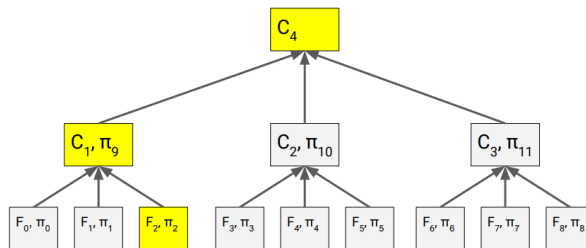


Рис.4. Доказательство в дереве Веркла с $l=3$ для листа со значением F_2

1. Производится проверка доказательства криптографического обязательства от F_2 до $C_1: \pi_2$.
2. Производится проверка доказательства криптографического обязательства от C_1 до $C_4: \pi_9$.

Для построения доказательства Веркла сестринские узлы не требуются, в отличие от доказательств в деревьях Меркла, а само доказательство по сути представляет из себя путь от узла, целостность которого необходимо проверить, до корня дерева.

По сравнению с доказательством в бинарном дереве Меркла, размер доказательства Веркла уменьшается с $O(\log_2 n)$ до $O(l \log_2 n)$:

$$\log_2 n = \frac{\log_2 n}{\log_2 l}$$

Недостатком деревьев Веркла, использующих векторные схемы обязательств, является то, что вычислительная стоимость и время построения и обновления таких деревьев Веркла больше вычислительной стоимости и времени построения и обнов-

ления бинарных деревьев Меркла. Уменьшение размера доказательства происходит за счет увеличения вычислительной стоимости (количества вычислений в l раз) для построения и обновления дерева: выбор коэффициента ветвления l определяет компромисс между вычислительной стоимостью и пропускной способностью. Так, если установить $l = 1024$, будет достигнуто 10-кратное уменьшение размера доказательства и, следовательно, пропускной способности за счет увеличения вычислительной стоимости в 1024 раза [3]. Если требуется меньшая вычислительная стоимость, l может быть изменена. Точное значение параметра l может варьироваться от приложения к приложению в зависимости от конкретных нужд. Ключевое преимущество деревьев Веркла в сравнении с бинарными, либо l -ичными деревьями Меркла заключается в их эффективности построения доказательства в случае большого числа блоков в дереве. Так, например, если дерево будет содержать ~миллиард блоков данных, доказательство в классическом бинарном дереве Меркла составит около 1 килобайта, тогда как в дереве Веркла доказательство составит менее 150 байт.

Преимущества деревьев Веркла

Деревья Веркла уменьшают размер доказательства членства, сокращая расстояние между листьями дерева и его корнем, а также устраняя необходимость предоставления близких узлов для проверки корневого хеша.

Еще большая экономия пространства достигается за счет использования схем полиномиальных обязательств вместо схем векторных обязательств. Полиномиальные обязательства позволяют доказательству членства иметь фиксированный размер независимо от количества листьев, которое оно подтверждает [4].

Перспективы применения деревьев Веркла

Совмещение схем полиномиальных обязательств (таких, как например: обязательство Педерсена и обязательство KZG) с деревом Веркла позволяет создать метод, который, будучи примененным в технологии блокчейн, обеспечит значительную оптимизацию всей системы. Пользователям таких распределенных реестров не придется хранить локальную копию базы данных о состоянии системы для проверки входящих блоков, как это сейчас организовано в большинстве решений на основе деревьев Меркла (в т. ч. деревьев Меркла Патриция), вызывая необходимость доступа пользователя к большому дисковому пространству. Вместо этого пользователям предлагается хранить «свидетельство» данных состояния, которое будет присылаться вместе с входящим блоком. Под свидетельством подразумевается комбинация коллекции отдельных фрагментов данных о состоянии, необходимых для выполнения определенного набора транзакций, и криптографического доказательства того, что свидетельство действительно является частью полных данных. Таким образом, отпадает необходимость хранить целостное состояние системы на жестком диске. Все, что будет необходимо для проверки входящего блока, будет уже

находиться в самом блоке. Такой подход, получивший название «клиенты без состояния» (от англ. «Statelessness»), обеспечит пользователям с мало-мощным оборудованием и небольшим дисковым пространством возможность оптимального взаимодействия с системами распределенных реестров [5].

На основе метода дерева Веркла могут быть также разработаны безопасные протоколы проверки целостности больших объемов данных, которые в ряде случаев будут являться менее трудоемкими по сравнению с имеющимися на сегодняшний день. Так, например, для решения задачи контроля целостности для удаленных хранилищ могут быть разработаны схемы аутентифицированного древовидного хеширования на основе метода дерева Веркла, в которых размер выходного хэш-кода будет иметь меньшие размеры.

Особое внимание уделяется постквантовой криптографии и разработке систем шифрования, способных обеспечить надежную защиту от квантовых атак. Квантовые компьютеры способны успешно осуществлять атаки на традиционные криптосистемы, которые широко используются на практике. Конкретно, системы, основанные на задаче факторизации целых чисел, такие как RSA.

Одними из предлагаемых методов, способных противостоять атакам квантовых компьютеров, являются постквантовые схемы электронной подписи на основе криптографических хеш-функции. Такие методы цифровой подписи считаются безопасными, поскольку используемые ими алгоритмы хеширования имеют низкую частоту возникновения коллизий. Подпись Меркла - один из таких алгоритмов. Проблема данных схем — очень большой размер подписи. В случае Подписи Меркла возникает проблема обхода дерева, связанная с нахождением эффективного алгоритма вычисления аутентификационных данных.

Существующая модель деревьев Веркла на основе векторных схем обязательств может предложить более эффективные процедуры проверки, давая возможность сохранять только необходимую информацию, что приведет к значительному уменьшению размеров цифровой подписи. Данный подход был предложен М. Явичем, Т. Кучухидзе и Т. Охрименко в своей научной работе «Verkle Tree-based Post-Quantum Digital Signature Scheme using Stateless Updatable Vector Commitment» [6].

Заключение

В данной работе было проведено сравнение деревьев Меркла с деревьями Веркла, выделена их роль в современном состоянии технологии «блокчейн». Были рассмотрены принципы построения обеих криптографических структур, выявлены их преимущества и недостатки. Установлено, что для задач проверки целостности больших объемов данных использование деревьев Веркла предпочтительнее классического бинарного дерева Меркла и l -ичных древовидных структур на его основе за счет достижения многократного преимущества в быстродействии и размере доказательства, в особенности при

использовании схем полиномиальных обязательств в случае с деревьями Веркла.

Практическое применение структуры деревьев Веркла имеет выдающиеся перспективы, которые начинают реализовываться уже сегодня.

Список литературы:

1. Миргасимов Д. Р. Нивелирование рисков использования технологии блокчейн на рынке финансирования устойчивого развития // Вестник евразийской науки. — 2023. — Т. 15. — № 6 — С. 2. URL: <https://esj.today/PDF/12ECVN623.pdf?ysclid=ixi4acusrw968716507>
2. Банк России. Цифровой рубль. Доклад для общественных консультаций [Electronic resource]. URL: https://cbr.ru/Content/Document/File/120075/concept_08042021.pdf
3. John Kuszmaul. Verkle Trees // Department of Mathematics Massachusetts Institute of Technology. - 2018.
4. Verkle trie for Eth1 state [Electronic resource]. URL: <https://dankradfeist.de/ethereum/2021/06/18/verkle-trie-for-eth1.html>
5. Деревья Веркла [Electronic resource]. URL: <https://ethereum.org/ru/roadmap/verkle-trees#current-progress>
6. Maksim Iavich, Tamari Kuchukhidz, Razvan Bocu. A Post-Quantum Digital Signature Using Verkle Trees and Lattices // Symmetry. — 2023. — P. 15(12):2165 — URL: https://www.researchgate.net/publication/376280585_A_Post-Quantum_Digital_Signature_Using_Verkle_Trees_and_Lattices

МЕТОДЫ ОБОДА СИСТЕМ ОБНАРУЖЕНИЯ/ПРЕДОТВРАЩЕНИЯ ВТОРЖЕНИЙ И СПОСОБЫ ЗАЩИТЫ ОТ НИХ

Сахаров К.Е.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе рассматриваются способы обхода систем обнаружения вторжений и систем предотвращения вторжений в контексте современных веб-приложений, а также способы защиты от них.

Введение

С постоянно растущим количеством кибератак возникает необходимость защищать подключенные к сети устройства. Одним из распространенных способов защиты являются системы обнаружения/предотвращения вторжений. Тем не менее такие системы зачастую уязвимы к различным техникам обхода. Обфускация - преобразование, применяемое к вредоносной полезной нагрузке, позволяющее пе-

редать данные без срабатывания сигнатурных сенсоров. Эти преобразования на различных уровнях OSI часто используется для обхода систем защиты [1], [2], [3]. В данной статье будут рассмотрены некоторые обфускации прикладного уровня OSI, которые могут быть использованы при атаках на веб-приложения, а также потенциальные способы защиты от обфусцированных атак.

Варианты обфускаций

Пример №1, применение URL Encode.

Для передачи небольшого количества данных веб-приложениям иногда используются GET параметры. Некоторые символы (например '?' и т.д.) в таких параметрах необходимо кодировать с помощью URL Encoding, но этим способом можно закодировать любой символ, указав его код последовательностью %XX, где XX - код байта в шестнадцатеричном формате. Например, код %41 равен символу "A". Также допустимы многобайтовые кодировки. Таким образом возникает возможность обфускации передаваемых данных.

Дальнейшие рассуждения приводятся на примере правила для системы IDS SNORT:

```
alert tcp any any -> $HOME_NET any  
(msg:"Generic Multi-OS x86 Shell via 'bin/sh' var 2  
(TCP protocol)";flow:established,to_client;content:"  
bin/sh"
```

Правило осуществляет поиск строки /bin/sh в TCP трафике, направленном от клиента к серверу.

Рассмотрим гипотетическую атаку инъекции команд на некоторое веб-приложение. Для проведения атаки необходимо отправить GET запрос к REST API приложения с параметром, содержащим инъекцию (для срабатывания правила искусственно обернем команду в "/bin/sh -c").

URL атаки: "http://<host>/api/vulnerable?x=`bin/sh -c id`". Отправка такого запроса вызовет срабатывание ранее рассмотренного правила IDS.

Однако, после применения простейшего URL Encode по всем символам к параметру запроса правило уже не детектирует его. URL принимает вид: "http://<host>/api/vulnerable?x=%60%2f%62%69%6e%2f%73%68%20%2d%63%20%69%64%60".

Пример №2, применение юникод последовательностей в строках JSON.

Не менее распространенным явлением в современных веб-приложениях, является использование формата сериализации JSON для входных и выходных данных API. Одним из типов данных, поддерживаемых JSON, являются строки. Для поддержки нестандартных символов юникода JSON строки поддерживают специальные экранирующие последовательности "\uXXXX", где XXXX - двухбайтный код символа шестнадцатеричном формате в соответствии с используемой кодировкой, но такое кодирование применимо к любым символам, что позволяет обфусцировать строки.

Теперь рассмотрим ту же атаку, но с REST API, принимающим JSON в теле POST запроса.

Для проведения атаки отправим POST запрос на "http://<host>/api/vulnerable_json" с содержимым

`{"x": "\/bin/sh -c id"}`). Такой запрос детектируется приведенным ранее правилом.

После применения описанной ранее обфускации строк JSON тело запроса примет вид `{"x": "\u0060\u002f\u0062\u0069\u006e\u002f\u0073\u0068\u0020\u002d\u0063\u0020\u0069\u0064\u0060"}`. Данная примитивная обфускация обходит защиту.

Пример №3, применение кодировки Base64.

Иногда с REST API также применяется дополнительная кодировка данных BASE64: данные кодируются на клиентской стороне и передаются в закодированном виде, затем декодируются на серверной стороне перед передачей в основную часть бизнес-логики приложения. Чаще всего такой подход используется для передачи произвольных двоичных данных в текстовом контейнере (таком как JSON, XML и т.д.), именно для этого была создана данная кодировка.

Рассмотрим на том же примере, но уже с предположением, что API принимает закодированные в BASE64 данные в JSON контейнере. После кодировки строки `\/bin/sh -c id` примет вид `YC9iaW4vc2ggLWMgaWRg`, а тело запроса – `{"x_b64": "YC9iaW4vc2ggLWMgaWRg"}`. Приведенная ранее сигнатура IDS пропускает такой запрос.

Пример №4, применение обфускации, специфичной для атаки

Теперь рассмотрим пример реальной атаки, использующей уязвимость (CVE-2021-44228, Log4Shell [4]) в компоненте Log4J - часто используемой библиотеке логирования для Java приложений, позволяющую выполнить произвольный Java код с помощью специально сформированного журнального сообщения. Некоторые элементы запросов, получаемых серверами, зачастую логируются, что делает эту уязвимость доступной извне и применимой к рассматриваемой области.

Сообщения логов, эксплуатирующие данную уязвимость, используют подстановочное выражение Log4J `"jndi"`: в простейшем случае сообщение будет содержать строку `/${jndi:ldap://example.com/payload}`, где `"example.com/payload"` - адрес сервера LDAP и путь до полезной нагрузки на нем. Одним из ранних механизмов защиты были простые сигнатуры IDS/IPS, осуществляющие поиск строки `/${jndi}`. Но такая защита очень легко обходится путем обфускации запроса с использованием других подстановочных выражений (например `"lower"`, которое не является опасным и может использоваться легитимно - заменяет все заглавные буквы в указанном выражении на строчные): `/${lower:jndi:ldap://example.com/payload}`. В сети Интернет быстро появилось большое количество скриптов, генерирующих более сложные обфусцированные строки.

Защита

Для детектирования обфусцированного подобным образом трафика модифицировать сигнатуры нецелесообразно, так как множество вариантов выходной строки огромно даже для небольшого числа представленных ранее примитивных обфускаций и

детектирование обфускаций должно быть добавлено в каждую сигнатуру системы, увеличивая размеры их баз. В качестве примера рассмотрим сигнатуру на строку `\/bin/sh` в чистом виде или полностью закодированную с URL Encode (игнорируя регистр).

`alert tcp any any -> $HOME_NET any (msg: "Generic Multi-OS x86 Shell via 'bin/sh' var 2 (TCP protocol)";flow:established,to_client;content:"/bin/sh";pcre:"/%2f%62%69%6e%2f%73%68/i"`

Строка `%2f%62%69%6e%2f%73%68` (буква `"h"` в конце не закодирована) не будет обнаружена данной сигнатурой.

В качестве защиты от обфусцированных атак можно применять контекстуальный анализ перед применением сигнатур [2]: так при обработке системой защиты URL адреса, произвести декодирование URL Encode перед применением сигнатур, то же самое со строками в JSON и т.д. Такой универсальный метод защиты тем не менее неприменим в случае обфускаций, используемых только в конкретных неизвестных ранее атаках (например, ранее рассмотренный Log4Shell). Еще одним способом защититься от данного вида обфускаций является установка детектора внутри процесса веб-сервера, что позволит проводить анализ получаемых извне данных после применения к ним преобразований уровня HTTP (например, URLDecode) - файрволл веб-приложений (WAF) [5] или даже после декодирования в коде самого приложения, но перед их передачей основной бизнес-логике. Также, в [6] предлагается выполнение обфусцированного кода (когда применимо) в изолированной среде с анализом поведения. Относительно новым успешно применяемым подходом для детектирования обфусцированных атак является использование классификаторов машинного и глубинного обучения [7]-[9].

Заключение

В ходе работы были рассмотрены примеры обфускаций трафика, применимых в контексте современных веб-приложений, которые можно использовать для эффективного обхода сигнатурного анализа, использующегося в большинстве систем обнаружения/предотвращения вторжений.

Для защиты от обхода была показана необходимость расширения сигнатурного анализа дополнительными мерами, такими как декодирование данных в зависимости от контекста, установка детектора внутри процесса веб-сервера, поведенческий анализ в песочнице или детекторы на основе методов машинного обучения.

Список литературы

1. Gorton A., Champion T. Combining evasion techniques to avoid network intrusion detection systems //North Chelmsford, MA, US: Skaion. – 2004.
2. Cheng T. H. et al. Evasion techniques: Sneaking through your intrusion detection/prevention systems //IEEE Communications Surveys & Tutorials. – 2011. – Т. 14. – №. 4. – С. 1011-1020.
3. Dyer K. P., Coull S. E., Shrimpton T. Marionette: A programmable network traffic obfuscation system

//24th USENIX Security Symposium (USENIX Security 15). – 2015. – С. 367-382.

4. CVE - CVE-2021-44228 [Электронный ресурс]. – URL: <https://cve.mitre.org/cgi-bin/cvename.cgi?name=cve-2021-44228>

5. Hacker A. J., CISSP I. Importance of web application firewall technology for protecting web-based resources //ICSA Labs an Independent Verizon Business. – 2008. – С. 7.

6. Andersson S., Clark A., Mohay G. Detecting network-based obfuscated code injection attacks using sandboxing //AusCERT Asia Pacific Information Technology Security Conference-AusCERT 2005. – AusCert Asia Pacific, 2005. – С. 13-25.

7. Homoliak I. et al. Improving network intrusion detection classifiers by non-payload-based exploit-independent obfuscations: An adversarial approach //arXiv preprint arXiv:1805.02684. – 2018.

8. Kim A., Park M., Lee D. H. AI-IDS: Application of deep learning to real-time Web intrusion detection //IEEE Access. – 2020. – Т. 8. – С. 70245-70261.

9. Homoliak I. et al. NBA of obfuscated network vulnerabilities' exploitation hidden into HTTPS traffic //The 9th International Conference for Internet Technology and Secured Transactions (ICITST-2014). – IEEE, 2014. – С. 310-317.

ЦЕНТР УПРАВЛЕНИЯ И МОНИТОРИНГА КОМПОНЕНТАМИ РАДИОЛОГИЧЕСКОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Краюшкин Д.В.

*Национальный исследовательский университет
«Высшая школа экономики»,
кафедра компьютерной безопасности
МИЭМ НИУ ВШЭ*

Аннотация

В работе представлено веб-приложение единого управления компонентами медицинской информационной системы, которая направлена на работу с радиологическими исследованиями и содержит в себе различные инструменты настройки, мониторинга и аналитики. Данное приложение позволяет управлять всеми сервисами, приложениями и приборами, входящих в систему PACS/RIS.

Введение

Современные информационные системы содержат в себе множество различных взаимосвязанных компонентов и узлов, которыми необходимо управлять и отслеживать их состояние. Медицинские информационные системы для работы с радиологическими изображениями (далее PACS/RIS - Picture archiving and communication system / Radiology Information System) достаточно сложно устроены и включают в себя не только сервисы по хранению и обработке информации, но и конфигурирование различных радиологических приборов или принтеров [1], а также средства аналитики и системы контроля

доступа. При увеличении нагрузки обслуживания, а также при росте количества медицинских учреждений, привязанных к RIS, использование сторонних программ и сервисов перестает быть эффективным и оптимизированным, поэтому требуется создание собственной системы управления данными, отвечающей задачам обслуживания и масштабирования.

В рамках работы представлены техническая реализация приложения единой системы управления компонентами PACS/RIS. Существующие решения на рынке не позволяют достаточно эффективно и быстро управлять компонентами большой системой. Как правило, такие решения разрабатываются под распространенные задачи диагностики конкретных компонентов больничной системы, тем самым удобство системы сводится к количеству инструментов, предоставляемыми разработчиками цифровых медицинских система. Кроме того, для работы используются вспомогательные средства и приложения, которые осуществляют операции с сервисами и службами операционной системой, где развернута цифровая медицинская система. В результате, администрирование системы является сложной и трудоемкой задачей. Разработанный комплекс предоставляет возможность гибкого управления медицинской информационной системой, направленной на создания, передачу, преобразования и хранения радиологических изображений.

Архитектура системы

На основании опыта взаимодействия с радиологической медицинской системой было произведено разделение основных используемых компонентов PACS/RIS в группы. На рис.1 представлена схема такой системы.



Рис.1. Схема групп компонентов, входящих в единую систему управления компонентами PACS/RIS

1. Объекты DICOM. Это группа компонентов включает средства управления и мониторинга компонентов, участвующих в формировании и обмене изображений формата DICOM, таких как принтеры, источники радиологических исследований, путей хранения изображений исследований.

2. Сервисы и приложения. Группа компонентов, которая включает в себя средства управления и мониторинга сервисов и приложений, запущенных в медицинской информационной системе. Например, сервис пересылки данных из RIS в ЕМИАС или сервис сжатия радиологических исследований.

3. Контроль доступа. Группа компонентов, которые осуществляет управление пользователями системы и определяет уровень привилегий.

4. Внешний и локальный обмен информацией. Группа компонентов содержит в себе средства отслеживания событий, которые происходит внутри или с другими внешними сервисами при обмене информацией. Например, предоставление хронологической истории отправки исследований с источника DICOM.

Информационная защита приложения

Важная роль при разработке уделена безопасному функционированию приложения. Для устранения возможности прямых атак приложение размещено в закрытом контуре, доступ к которому осуществляется через защищенную сеть, сертификаты для которой постоянно обновляются.

Обмен данными между сервером и клиентов осуществляется на основе доверенных сертификатов по протоколу HTTPS. Причем проверка подлинности сертификата происходит как со стороны серверных сервисов, так и со стороны клиентского приложения. Таким образом, обмен данными возможен только для защищенных соединений. Для дополнительного мониторинга безопасности, в том числе проверка сертификатов, серверный контур обслуживают лицензированные программные решения по контролю трафика, такие как Kaspersky Security Center. В ее задачи входит блокирование всей подозрительной активности.

Приложение реализовано на базе WebAssembly, которое осуществляет исполнение кода на стороне клиента, тем самым являясь высокопроизводительным и аппаратно-независимым. Однако WebAssembly является потенциально уязвимым для атак по вспомогательным каналам на аппаратном уровне. Для обеспечения защиты приложения используются доступные средства браузера и платформы ASP.NET Core. Сюда входят: блокирование системой безопасности запросы к другим доменам (cross-origin resource sharing *сокр.* - CORS); предотвращение атак с межсайтовой подделкой запросов (cross-site request forgery *сокр.* - CSRF). Кроме того, код приложения разрабатывался таким образом, чтобы учесть рекомендации для запрета межсайтовых сценариев (cross-site scripting *сокр.* - XSS) [2]

Для проверки подлинности приложения и пользователей были выбраны протокол на основе токенов (JSON Web Tokens *сокр.*- JWTs). Это связано с тем, что уменьшается контактная зона атак; API и конечные точки сервера защищены от подделки маркеров межсайтовых запросов; возможность клиента и сервера верифицировать измененный токен; короткий срок жизни токена.

Для определения прав пользователя и доступа к ресурсам на их основе в серверном контуре предусмотрен единый сервис авторизации с стандартизированным протоколом аутентификации OAuth 2.0 и его реализации OpenID Connect в ASP.NET Core.

Контроль доступа

Контроль доступа осуществляется на основе единого сервера авторизации, в основе которого лежит протокол OAuth 2.0. Сервис авторизации выдает пользователю токен доступа (JWT токен), на основании которого он может получить доступ к тем или иным ресурсам в зависимости от уровня доступа. Преимущества протокола и его реализаций состоит в едином сервисе авторизации, который определяет уровень доступа для всех ваших веб-приложений, сервисов и баз данных. Кроме того, такой подход позволяет использовать технологию одного входа (single sign on), когда нет необходимости прохождения форм авторизации для каждого ресурса или приложения. Авторизация в этом случае происходит, например, при входе пользователя в защищенную сеть.

Этап аутентификации OAuth 2.0 по типу Authorization flow with PKCE (Authorization Code Flow with Proof Key for Code Exchange *непр.*-Поток кода авторизации с ключом подтверждения для обмена кодами) включает в себя следующий набор действий и обмен с сервером авторизации: браузер, в котором открыто клиентское приложение создает криптографически-случайный код (code verifier), на основе него генерируется код вызова (code challenge). Последний вместе с методом запроса авторизации(/authorize) отправляется на сервер авторизации. Пользователь получает страницу входа и авторизации, проходит аутентификацию и получает код авторизации, позволяющий получать токены. Далее отправляет код авторизации и code verifier вместе с методом запроса токенов (/oauth/token) отправляется на сервер авторизации. Сервер авторизации проверяет совпадение code verifier и code challenge и возвращает запрашиваемые токены (ID токен и токен доступа).

В настоящей работе для реализации сервиса авторизации протокола OAuth 2.0 на платформе ASP.NET core был использован программный продукт с открытым исходным кодом OpenIdDict [2].

Техническая реализация

Единая система управления компонентами PACS/RIS представляет из себя клиент-серверное приложение, которое развернуто в той же среде, что и набор средств PACS. На рис.2 представлено схема наполнения веб-приложения, где в блоках перечислены разделы, которые представлены как наборы веб-страниц.



Рис.2. Разделы системы управления компонентами PACS/RIS

Разделы «Источники DICOM», «Адресаты DICOM» и «Принтеры DICOM» осуществляет конфигурирование источников радиологических изображений и принтеров, которые размещены в медицинском учреждении. Для источников устанавливаются такие параметры как модальность, отделение, комнаты, а также производятся сетевые настройки. В случае принтеров предусмотрена возможность установки режимов печати.

Раздел «Буфер DICOM» представляет информацию об сообщениях, которые содержат исследования с изображениями. В сообщения формата DICOM входят не только изображения, но и вспомогательная данные, которые описывают детали и параметры исследование, а также информацию о пациенте. Данный раздел представляет инструменты для удобного визуального отображения.

Раздел «Пути хранения» и «Статистика путей» позволяет управлять локальными путями хранения исследований, которые отправляет источник радиологических изображений. Исследования сохраняются локально до тех пор, пока сервис упаковки не отправит их или на общий сервер PACS/RIS, или в MIS и т.д. Практический опыт показал, что при ограниченных ресурсах памяти необходимо автоматизировано мониторить эти директории, в том числе для выявления «застрявших» исследований. Вследствие этого, был разработан сервис и раздел, осуществляющий сбор информации о состоянии путей хранения.

Раздел «Журнал событий» представляет из себя инструмент для отображения событий, которые фиксируют различные сервисы и приложения. Внутри раздел оснащен большим количеством фильтров,

возможностью динамических переходов для отслеживания причин каких-либо событий и происшествий. На рис.3 представлен пример модального окна с деталями обработки радиологического исследования одним из сервисов.

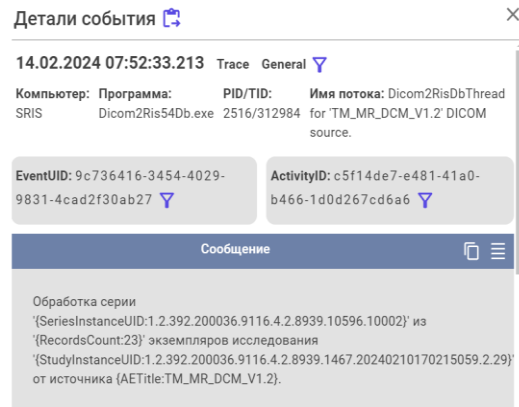


Рис.3. Модальное окно деталей событий

Раздел «Пользователи» осуществляет управление пользователями и их доступом.

Раздел «Сервисный менеджер» содержит инструменты управления и информацию о запущенных сервисах и приложениях, развернутых в функционирующей операционной системе.

Веб-приложение реализована с использованием технологии Blazor платформы dotNET [4]. Поддерживает работу с базами данных MySQL и PostgreSQL.

Заключение

Практический опыт использования и масштабирования медицинской системы обработки, хранения и передачи радиологических изображений PACS/RIS показал необходимость единого центра управления. Было разработано веб-приложение, которое обеспечивает легкий доступ, мониторинг и управления большим количеством компонентов системы.

Разработанное приложение проходит промышленную эксплуатацию в локальных PACS/RIS различных коммерческих и государственных медицинских учреждениях. Компоненты продолжают добавляться при увеличении нагрузки или сложности их использования.

Список литературы

1. А. В. Гаврилов, И. В. Архипов, И. В. Куликов, А. М. Ятченко Цифровые технологии при работе с медицинскими изображениями // Вестник рентгенологии и радиологии. – 2010. – № 4. – С. 057-063
2. Prevent Cross-Site Scripting (XSS) in ASP.NET Core. <https://learn.microsoft.com/en-us/aspnet/core/security/cross-site-scripting> (дата обращения: 31.05.2024)
3. OpenIdDict URL: <https://documentation.openidict.com/index.html> (дата обращения: 14.02.2024)
4. Blazor URL: <https://dotnet.microsoft.com/en-us/apps/aspnet/web-apps/blazor> (дата обращения: 14.02.2024)

ИССЛЕДОВАНИЕ УГРОЗ И РИСКОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ФИНАНСОВО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ НИУ ВШЭ

Гришкина В.И.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Основной целью данного исследования является анализ и оценка угроз и рисков информационной безопасности финансово-аналитической системы НИУ ВШЭ. В работе проводится всесторонний анализ:

- Выявление и классификация потенциальных источников угроз (внешних и внутренних).
- Оценка вероятности реализации различных типов угроз (нарушение конфиденциальности, целостности и доступности информации).
- Расчет и ранжирование рисков информационной безопасности с учетом вероятности и возможного ущерба.

Актуальность исследования обусловлена ключевым значением этой системы для деятельности университета. Она обрабатывает и хранит критически важную финансовую и управленческую информацию, нарушение доступности, целостности или конфиденциальности данных, в которой может иметь серьезные последствия. Кроме того, система интегрирована с различными внешними системами, увеличивая поверхность атаки, а также содержит конфиденциальную информацию, утечка которой может повлечь репутационные и правовые риски. Финансово-аналитическая система является критически важным элементом ИТ-инфраструктуры университета, от нее зависит непрерывность основных бизнес-процессов.

Введение

Обеспечение информационной безопасности бизнес-систем является важным аспектом в современном мире, где цифровые технологии играют ключевую роль во всех аспектах деятельности компаний. В условиях все возрастающих киберугроз и утечек информации, обеспечение безопасности информации становится приоритетной задачей для любого предприятия. Это означает, что бизнесам необходимо принимать меры для защиты своих информационных активов, включая конфиденциальные данные клиентов, финансовую информацию, интеллектуальную собственность и другие чувствительные ресурсы. [1]

Оценка рисков и угроз является основой для разработки эффективных стратегий и мер по обеспечению безопасности. Ключевыми шагами, которые необходимо включить в процесс оценки рисков и угроз являются:

Идентификация активов: определение всех информационных активов, систем, данных и ресурсов, которые необходимо защищать.

Определение угроз: анализ потенциальных угроз, которые могут повлиять на безопасность активов. Включает в себя угрозы кибератак, физических вторжений, природных бедствий и т.д.

Оценка уязвимостей: определение уязвимостей в системах и процессах, которые могут быть использованы злоумышленниками для атаки на активы.

Оценка рисков: оценка вероятности возникновения угроз и величины ущерба, который они могут причинить [2].

Финансово-аналитическая система НИУ ВШЭ

Финансово-аналитическая система НИУ ВШЭ – это специализированная информационная система на базе 1С предприятия, предназначенная для учета, анализа и планирования финансовой деятельности университета.

ФАС предоставляет широкий функционал, который включает в себя следующие возможности:

1. Учёт и анализ финансовых операций: система производит учёт и анализ всех финансовых операций университета, включая доходы, расходы, бюджетирование, учёт средств и т.д.

2. Бюджетирование и планирование: система поддерживает процессы бюджетирования и планирования финансовых показателей университета, включая составление бюджетов, утверждение бюджетов, мониторинг исполнения бюджетов и анализ отклонений.

3. Финансовая отчётность: система предоставляет возможность формирования различных финансовых и аналитических отчётов, финансовых показателей и дашбордов для управленческого анализа и принятия решений.

4. Управление финансовыми рисками: система может включать инструменты для управления финансовыми рисками, анализа финансовой устойчивости и прогнозирования финансовых показателей.

5. Интеграция с другими системами: система обеспечивает возможность интеграции с другими информационными системами университета, такими как системы управления учебным процессом, кадровые системы и т.д.

6. Управление доступом и безопасностью: система обеспечивает управление доступом к финансовым данным в соответствии с ролями и полномочиями пользователей, а также обеспечивает безопасность и конфиденциальность финансовой информации. [3]

Информационные активы системы

1. Программное обеспечение: лицензионное программное обеспечение, используемое для управления финансовыми данными, аналитики, отчетности и других функций.

2. Аппаратное обеспечение: сервера, компьютеры, сетевое оборудование и другие устройства, необходимые для функционирования финансово-аналитической системы.

3. Базы данных: хранение и управление финансовыми данными и информацией, которые используются в системе для обеспечения аналитики и отчетности.

4. Людские ресурсы: персонал, который управляет, обслуживает и поддерживает финансово-аналитическую систему, включая администраторов, аналитиков и других специалистов.

5. Лицензии и договора: документация, связанная с лицензиями на программное обеспечение, а также договора с поставщиками услуг и оборудования.

6. Процессы и процедуры: документированные процессы, процедуры и стандарты, которые регулируют использование и обслуживание финансово-аналитической системы. [4]

Определение угроз

Угрозы информационной безопасности системы — это потенциальные события или действия, которые могут нанести ущерб конфиденциальности, целостности или доступности информации, а также функционированию информационной системы в целом. Угрозы могут происходить как из внутренних источников (например, неосторожное поведение сотрудников), так и из внешних (например, злонамеренные кибератаки).

К основным внутренним угрозам ФАС можно отнести:

1. Несанкционированный доступ к информации: возможность несанкционированного доступа к финансовым данным сотрудниками, студентами или другими пользователями системы.

2. Несанкционированное изменение данных: возможность манипуляции финансовыми данными или отчетами внутри системы, что может привести к искажению финансовой информации.

3. Недостатки в управлении доступом: недостаточное управление доступом к финансовой информации, неправильная настройка прав доступа и слабые пароли.

4. Халатность сотрудников: случайное действие сотрудников, которое может привести к потере или искажению информации.

5. Подкуп и вербовка: диверсия, шпионаж, шантаж, саботаж.

6. Разглашение коммерческой тайны.

7. Физическое хищение аппаратных средств.

8. Системные сбои, которые могут происходить по различным причинам, включая технические сбои оборудования, программные ошибки, кибератаки, неправильные настройки системы, человеческие ошибки. [5]

Для определения основных внешних угроз безопасности рассмотрим методологию MITRE ATT&CK, которая является ведущим инструментом для обнаружения и реагирования на кибератаки. Система была разработана компанией MITRE Corporation с целью создания единой методологии для анализа и оценки кибератак.

MITRE ATT&CK — это система, которая одновременно представляет собой базу знаний о типичных действиях при проведении кибератак и инстру-

мент для обнаружения киберугроз. Она состоит из матрицы, где отражены типичные действия злоумышленников (tactics), и техник, используемых ими при атаке на корпоративные сети. [6]

Ниже приведен перечень некоторых техник атак, описанных в матрице MITRE ATT&CK, которые могут быть применены для финансово-аналитической системы НИУ ВШЭ:

1. Хищение учетных данных: Эта техника заключается в хищении учетных данных для получения доступа к системам компании. Она может быть осуществлена с помощью различных методов, таких как социальная инженерия, вирус или фишинг.

2. Внедрение вредоносного кода: программы, которые автоматически распространяются по сетям и заражают компьютеры. Они могут быть созданы для утечки данных, получения удаленного доступа и других целей.

3. Компрометация учетной записи: Эта техника заключается во взломе учетных записей пользователей, чтобы получить доступ к системам или данным компании.

4. Кража данных: может быть осуществлена с помощью программного обеспечения для незаконного доступа, проникновения через уязвимости в коде или физической кражи хранилища данных.

5. Аварийные остановки сервиса (DDoS-атака): техника заключается в использовании множества компьютеров для генерации большого количества трафика на конкретный сайт или сервис. Вызывает отказ в обслуживании и недоступность сайта или сервиса для пользователей.

6. Отказ в обслуживании может возникнуть из-за различных причин, таких как технические сбои, недоступность серверов, кибератаки, стихийные бедствия

Оценка рисков

После того, как угрозы и уязвимости идентифицированы, происходит оценка рисков, связанных с этими угрозами и уязвимостями.

Согласно ГОСТ Р ИСО/МЭК ТО [7] для расчета уровня рисков организации используется формула:

$$R = P(t) \times P(v) \times S,$$

где

R - уровень риска

$P(t)$ - вероятность реализации угрозы информационной безопасности. Величина варьируется от 1 до 3.

$P(v)$ - вероятность наличия уязвимости. Величина варьируется от 1 до 3.

S - ценность актива. Величина варьируется от 0 до 4.

В таблице 1 приведены расчеты рисков информационной безопасности для системы:

Таблица 1. Оценка рисков ФАС

| Угроза | $P(t)$ | $P(v)$ | S | R |
|---|--------|--------|-----|-----|
| 1. Несанкционированный доступ к информации | 2 | 3 | 4 | 24 |
| 2. Несанкционированное изменение данных | 1 | 3 | 4 | 24 |
| 3. Недостатки в управлении доступом | 2 | 2 | 4 | 16 |
| 4. Несанкционированный доступ к данным | 2 | 3 | 4 | 24 |
| 5. Халатность сотрудников | 3 | 3 | 1 | 9 |
| 6. Подкуп и вербовка | 2 | 2 | 2 | 8 |
| 7. Разглашение коммерческой тайны | 2 | 3 | 3 | 18 |
| 8. Физическое хищение аппаратных средств | 2 | 2 | 3 | 12 |
| 9. Системные сбои | 2 | 3 | 1 | 6 |
| 10. Хищение учетных данных | 2 | 2 | 3 | 12 |
| 11. Внедрение вредоносного кода | 3 | 3 | 3 | 27 |
| 12. Компрометация учетной записи | 2 | 2 | 4 | 16 |
| 13. Кража данных | 2 | 2 | 3 | 12 |
| 14. Аварийные остановки сервиса вследствие атак | 3 | 2 | 2 | 6 |
| 15. Отказ в обслуживании | 2 | 1 | 2 | 4 |

Для более наглядного представления рисков был проведен анализ рисков "Мастерпланс"[8]. На рис.1 представлена матрица рисков:



Рис.1. Матрица рисков

На рис.2 представлено распределение по степени риска:

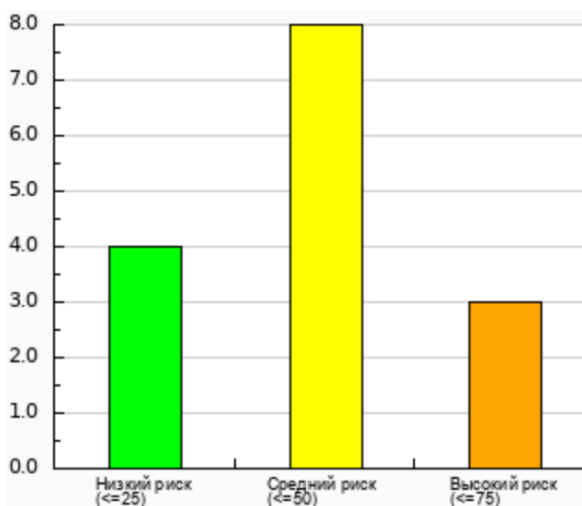


Рис.2. Степень риска

Заключение

В ходе исследования были выявлены различные риски и угрозы информационной безопасности финансово-аналитической системы НИУ ВШЭ. Эти риски включают в себя возможные атаки хакеров, утечки конфиденциальной информации, а также недостаточную защиту от внутренних угроз. Основываясь на полученных результатах оценки рисков, можно сделать следующие выводы:

Наиболее высокие риски связаны с угрозами:

1. Внедрение вредоносного кода
2. Несанкционированный доступ к информации, несанкционированное изменение данных, несанкционированный доступ к данным.

Эти риски имеют высокие оценки по вероятности реализации и возможному ущербу, поэтому они являются приоритетными для принятия мер по снижению.

Средний уровень рисков связан с такими угрозами, как:

1. Разглашение коммерческой тайны.
2. Недостатки в управлении доступом, компрометация учетной записи.
3. Физическое хищение аппаратных средств, хищение учетных данных, кража данных.

Для этих рисков необходимо разработать и внедрить соответствующие меры защиты.

Сравнительно низкий уровень рисков связан с такими угрозами, как:

1. Халатность сотрудников.
2. Подкуп и вербовка.
3. Системные сбои, аварийные остановки сервиса вследствие атак.
4. Отказ в обслуживании.

Для этих рисков необходимо поддерживать существующие меры защиты и контроля.

Таким образом, результаты оценки рисков позволяют определить приоритетные направления для совершенствования системы информационной безопасности финансово-аналитической системы НИУ ВШЭ.

Для уменьшения этих рисков необходимо принимать меры по улучшению системы защиты информации, обучению сотрудников правилам информационной безопасности, а также регулярному аудиту системы на предмет обнаружения уязвимостей.

Список литературы

1. ГОСТ Р 57580.1-2017 Безопасность финансовых (банковских) операций Защита информации финансовых организаций
2. Manoj K. S. Cyber risk in banking services: the extent of cyber risks provisions and security measures //International Journal of Management (IJM). – 2021. – Т. 12. – №. 1. – С. 1332-1339. (Дата обращения 03.02.2024).
3. Справочники системы [Электронный ресурс]. – URL: <https://confluence.hse.ru/pages/viewpage.action?pageId=90355005> (Дата обращения 06.02.2024). – Текст: электронный.
4. Размещение данных 1С:Предприятия 8 [Электронный ресурс]. – URL:<https://1c-predpriyatie-qlik.ivan-shamaev.ru/1s-predpriyatie-8-opisanie-bazy-dannyh/> (Дата обращения 07.02.2024). – Текст: электронный.
5. Лекция 2. Основные угрозы безопасности предприятия. Учебные вопросы [Электронный ресурс]. – URL: <https://present5.com/lekciya-2-osnovnye-ugrozy-bezopasnosti-predpriyatiya-uchebnye-voprosy/> (дата обращения 05.02.2024).
6. MITRE АТТ&СК: [сайт]. – URL: <https://attack.mitre.org/> (Дата обращения 05.02.2024). – Текст: электронный.
7. ГОСТ Р ИСО/МЭК ТО 13335-3-2007
8. АНАЛИЗ РИСКА: [сайт]. – URL: <https://www.masterplans.ru!/risk/> (Дата обращения 08.02.2024). – Текст: электронный.

РАСПОЗНАВАНИЕ АУДИО КОНТЕНТА ПО ДАННЫМ АКСЕЛЕРОМЕТРА

Жидков А.С.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент больших данных
и информационного поиска
ФКН НИУ ВШЭ*

Аннотация

В работе рассматривается уязвимость, связанная с использованием данных акселерометра мобильного устройства для несанкционированного распознавания аудио контента, воспроизводимого как на самом устройстве, так и на близлежащих устройствах. Рассмотрена экспериментальная модель тракта “динамик – акселерометр”, предложен алгоритм распознавания аудио контента по данным акселерометра, разработан прототип мобильного приложения, реализующий предложенный алгоритм распознавания в реальном времени, проведен анализ качества распо-

знавания аудио контента для некоторых моделей мобильных устройств.

Введение

В современном смартфоне можно обнаружить десятки различных сенсоров, которые превращают телефон в «умное» устройство. Все сенсоры можно условно разделить на два типа: (а) «опасные» сенсоры, использование которых может нарушить приватность пользователя, (б) относительно безопасные сенсоры, которые, как правило, не нарушают приватность пользователя.

К сенсорам первого типа традиционно относятся камера и микрофон. Использование таких сенсоров жестко регламентируется специальными разрешениями на уровне операционной системы (ОС). Так, например, для получения доступа к камере или микрофону приложение должно запросить доступ в момент запуска приложения с помощью системного диалога, а при использовании микрофона в современных мобильных ОС (например, Андроид 12 и выше) появляется специальный индикатор, предупреждающий пользователя о том, что то или иное приложение ведет запись аудио.

С другой стороны, существует целый набор «безопасных» сенсоров, доступ к которым или никак не ограничен или предоставляется без уведомления пользователя. К таким сенсорам относятся, например, датчик освещенности, датчик приближения, гироскоп, акселерометр и ряд других сенсоров. Учитывая возросший интерес к обеспечению приватности пользователей исследование возможности несанкционированного использования “безопасных” датчиков является актуальной задачей. В рамках настоящего исследования нас интересует один из самых популярных и широко распространенных сенсоров – акселерометр. Практически любое современное мобильное устройство содержит акселерометр, который традиционно используется для того, чтобы: (а) определять ориентацию устройства в пространстве, (б) в качестве устройства управления в мобильных играх или (в) как датчик определения активности пользователя.

Несмотря на «безобидность» акселерометра есть основания полагать что он может быть использован не по своему прямому назначению. В ряде работ [1-4] была предпринята попытка использовать акселерометр для распознавания речи и было показано, что распознавание речи с его помощью если и возможно, то лишь частично (отдельные фонемы) и требует специфической настройки под конкретное устройство [3, 4]. В данной работе мы рассматриваем несколько иную угрозу, а именно распознавание с помощью акселерометра заранее известного аудиоконтента (как это делают, например, популярные приложения распознавания музыки: Shazam, SoundHound, MusicID и др.) при его проигрывании на самом устройстве или на другом устройстве, с которым непосредственно контактирует смартфон пользователя.

Модель тракта “динамик – акселерометр”

Анализ открытых баз данных, в частности, базы данных мобильного приложения PhyPhox университета RWTH Aachen [5] (на момент написания статьи в базе содержалась информация о 3702 моделях устройств и 31531 пользователях) показал, что на текущий момент около 99% всех мобильных устройств, находящихся в эксплуатации, имеют встроенный акселерометр. Распределение максимальной частоты дискретизации для устройств на базе ОС Андроид, построенное по данным собранным приложением PhyPhox показало, что около 55% устройств на базе ОС Андроид имеют возможность считывать данные акселерометра с частотой дискретизации (F_{acc}) от 400 до 500 Гц.

В работах [1-4] не проводится полноценный анализ модели тракта “динамик – акселерометр”, вместо этого делается общее предположение о том, что аудио сигнал приводит к появлению в спектре сигнала акселерометра большого количества “комплексных гармоник”, поэтому обычно предлагается сначала привести сигнал к единой частоте дискретизации и затем обучить нейронную сеть на большом количестве экспериментальных данных “исходный аудио сигнал” – “сигнал акселерометра”. Такой подход, с одной стороны, требует сбора большой обучающей выборки, а с другой – дает плохие результаты при использовании нейросети, обученной на другой модели телефона. В настоящей работе предпринята попытка построить общую модель типового тракта “динамик – акселерометр”. Для этого были записаны отклики акселерометра для двух типов аудио сигналов: (а) линейно-частотно модулированного (ЛЧМ) сигнала с частотой от 0 до 2000 Гц и скоростью модуляции 5 Гц/сек и (б) комбинированного сигнала, состоящего из такого же ЛЧМ сигнала и однонального сигнала с частотой 1000 Гц. Эксперименты были проведены на нескольких моделях смартфонов: Samsung A50, Samsung Galaxy S22 Ultra, Samsung Z Fold 5, Google Pixel 6, Google Pixel 7, Huawei Nova 11 и Honor 50. Пример спектрограммы сигнала акселерометра (ускорение по оси Z) для одного из смартфонов представлен на рис. 1.

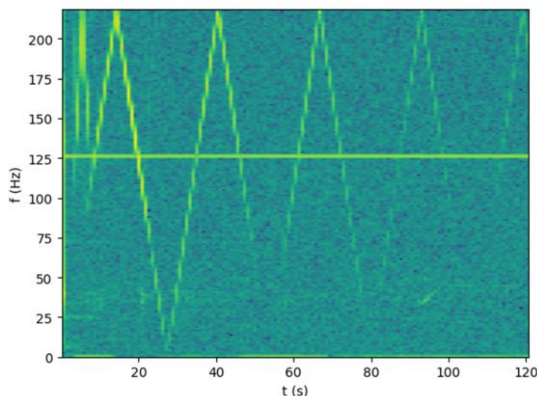


Рис. 1. Спектрограмма отклика акселерометра на комбо сигнал: ЛЧМ 0-2кГц + 1 кГц (Pixel 6)

Как показал проведенный анализ тракт “динамик-акселерометр” не вносит существенных нелинейных искажений. “Аналоговая” часть тракта относительно линейна и имеет достаточно широкий частотный диапазон, как минимум, до 1.5-3 кГц, в зависимости от модели устройства, а вся нелинейность обусловлена тем, что частота дискретизации относительно широкополосного входного сигнала в несколько раз ниже необходимой частоты Котельникова-Найквиста.

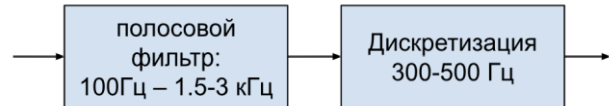


Рис. 2. Модель тракта “динамик-акселерометр”

Таким образом, обобщенная типовая модель тракта “динамик-акселерометр” выглядит для всех телефонов примерно так, как изображено на рис.2.

Алгоритм распознавания

Для формирования сигнатур контента по данным акселерометра было решено адаптировать алгоритмы, которые хорошо зарекомендовали себя при решении задач распознавания контента по звуку, см. например [6]. Ключевое отличие нашей задачи от задачи распознавания контента по звуку заключается в ограниченности частоты дискретизации акселерометра, а также в наличии отклика тракта передачи на частотах, которые в несколько раз превышают частоту дискретизации. Чтобы не терять информацию, которая содержится в спектре звукового сигнала в диапазоне 0.3-2кГц было решено не проводить фильтрацию и ресэмплирование сигнала акселерометра. Алгоритм формирования сигнатур как на стороне устройства, так и на этапе формирования “эталонных” сигнатур работает на максимальной поддерживаемой частоте дискретизации сигнала для конкретного устройства, поэтому первым этапом распознавания является автоматическая оценка частоты дискретизации сигнала акселерометра. Такой подход усложняет формирование базы “эталонных” сигнатур, так как, по существу, в данном случае требуется уникальная база “эталонных” сигнатур для каждой частоты дискретизации, которая может встретиться в реальных устройствах. Однако на практике это не представляет серьезных проблем: во-первых, база “эталонных” сигнатур формируется заранее и не требует вычислений в реальном времени, во-вторых, как показал анализ характеристик акселерометров, представленный в предыдущем разделе, большая часть устройств поддерживает частоту дискретизации в диапазоне 400-500Гц, причем ряд частот из этого диапазона встречается крайне редко, поэтому для поддержки большинства мобильных устройств достаточно подготовить несколько десятков вариантов базы “эталонных” сигнатур с точностью определения частоты дискретизации примерно 1-2 Гц. Таким образом, сигнатуры на стороне мобильного устройства (т.е. по данным акселерометра) формируются без

ресэмплирования и дополнительной фильтрации на максимальной рабочей частоте акселерометра F_{acc} , а для создания базы "эталонных" сигнатур из аудио файлов мы предварительно ресэмплируем исходный аудиосигнал до частоты kF_{acc} , где k – коэффициент передискретизации ($k > 20$), затем сигнал фильтруется с помощью цифрового фильтра с полосой пропускания 100Гц – 2кГц, чтобы эмулировать усредненную модель тракта "динамик-акселерометр" изображенную на рис.2, затем децимируется с коэффициентом 1:k и только затем подается на блок формирования сигнатур.

Для дальнейшего формирования сигнатур используется алгоритм Хатсима-Калкера [6] с небольшими модификациями. Алгоритм состоит из следующих этапов:

1. Разбиение непрерывного сигнала на частично перекрывающиеся сегменты длиной N отсчетов x_n , где $n = 0, \dots, N - 1$.

2. Применение к сегменту окна Хэмминга:

$$y_n = x_n (0.54 - 0.46 \cos(2\pi n/N))$$

3. Вычисление дискретного преобразования Фурье (ДПФ) сегмента:

$$X_k = \sum_{n=0}^{N-1} y_n \exp\left(\frac{2\pi kn}{N}\right), k = 0, \dots, N/2 - 1$$

4. Группировка квадратов значений ДПФ на несколько неперекрывающихся частотных суб-полос, в результате на выходе такого преобразования получаются значения $E(m, t)$ - энергий сигнала в m -й полосе для t -го временного сегмента.

5. Формирование по аналогии с [6] блока бинарных сигнатур:

$$F(m, t) = \begin{cases} 1 & \text{if } E(m, t) - E(m, t - D) - (E(m + 1, t) - E(m + 1, t - D)) > 0 \\ 0 & \text{if } E(m, t) - E(m, t - D) - (E(m + 1, t) - E(m + 1, t - D)) \leq 0 \end{cases}$$

где D – глубина задержки (для оригинального алгоритма [6] $D=1$, но эмпирическим путем мы определили, что оптимальные результаты распознавания получаются при $D \gg 1$).

На практике были подобраны следующие параметры алгоритма: размер сегмента $N = 128$, фактор перекрытия сегментов: 15/16, количество частотных полос: $m = 0, \dots, 24$ (т.е., количество бит на каждую суб-сигнатуру = 24), задержка: $D = 12$.

После формирования блока сигнатур акселерометра длительностью около 15 секунд проводится процедура нечеткого поиска фрагмента по базе "эталонных" сигнатур. Для этого указанный блок сигнатур по-сегментно "скользит" по сигнатурам "эталонных" записей и на каждом шаге такого "скольжения" вычисляется расстояние Хэмминга между блоком сигнатур акселерометра и текущим блоком "эталонных" сигнатур. В случае, если расстояние Хэмминга оказывается ниже некоторого порога считается, что обнаружено совпадение с соответствующим фрагментом "эталонной" базы аудио образцов.

Результаты

Указанный алгоритм был реализован в прототипе мобильного приложения Android на языке Kotlin. В качестве "эталонной" базы контента использовалась база из 228 разножанровых музыкальных композиций. Формирование сигнатур акселерометра, а также нечеткое сравнение сигнатур с "эталонной" базой сигнатур производилось непосредственно на мобильном устройстве. Тестовый сигнал представлял собой набор одноминутных нарезок из 113 случайно выбранных композиций (общая длительность 1 час 53 минуты). Порог распознавания был подобран таким образом, чтобы за 2 часа непрерывного распознавания ни на одном из телефонов не было зафиксировано ложных распознаваний контента из "эталонной" базы. Источником звука являлся ноутбук MacBook Air M1, на котором воспроизводился тестовый сигнал с громкостью, примерно 70% от максимально возможной громкости звука. Во время теста телефоны с установленным мобильным приложением помещались на корпус ноутбука рядом с тачпадом. Так как окно анализа составляло около 15 секунд за одну минуту длительности фрагмента тестовой композиции декодер мог выполнить несколько последовательных попыток распознавания. Композиция считалась корректно распознанной, если она правильно распознана хотя бы один раз за отведенный одноминутный интервал и при этом за данный интервал не было ложных распознаваний. Точность распознавания (т.е. доля правильно распознанных композиций) для нескольких протестированных моделей телефонов приведены в табл.1.

Таблица 1.

| Устройство | F_{acc} | Точность |
|-------------------|-----------|----------|
| Samsung S22 Ultra | 415 Гц | 95.6% |
| Samsung Z Fold 5 | 470 Гц | 89.4% |
| Google Pixel 6 | 435 Гц | 94.7% |
| Huawei Nova 11i | 400 Гц | 85.0% |

Заключение

Результаты работы подтвердили гипотезу о том, что акселерометр может использоваться для распознавания аудио контента, воспроизводимого на самом устройстве или в случае, если источник звука физически соприкасается с мобильным устройством (например, лежит с ним на одном столе). Результаты работы могут быть полезны для выработки рекомендаций по защите от подобных угроз, а также рекомендаций по обеспечению более гранулярного доступа к микрофону на мобильных устройствах.

Список литературы

1. S Abhishek Anand, et al. Spearphone: a lightweight speech privacy exploit via accelerometer-sensed reverberations from smartphone loudspeakers // Proceedings of the 14th ACM Conference on Security

and Privacy in Wireless and Mobile Networks. WiSec '21. Abu Dhabi, UAE, июнь 2021, с.288-299

2. Roberto De Prisco, et al. An improved privacy attack on smartphones exploiting the accelerometer // Journal of Information Security and Applications, том. 75, июнь 2023, с.1-17

3. Pengfei Hu, et al. AccEar: Accelerometer Acoustic Eavesdropping with Unconstrained Vocabulary // 2022 IEEE Symposium on Security and Privacy (SP). IEEE, май 2022, с.1757-1773

4. Weigao Su, et al. Towards Device Independent Eavesdropping on Telephone Conversations with Built-in Accelerometer // Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5(4):1-29, дек. 2021, с.177:1-177:29

5. RWTH aachen university. phyphox Sensor Database. URL: <https://phyphox.org/sensordb/> (дата обращения 18.10.2023)

6. J.A. Haitzma, A.A.C.M. Kalker. A highly robust audio fingerprinting system with an efficient search strategy // Journal of New Music Research 32.2 (2003), с.211-222

УЯЗВИМОСТИ ИНФОРМАЦИИ ПРИ ПЕРЕДАЧЕ АВТОИНКРЕМЕНТНЫХ ИДЕНТИФИКАТОРОВ ПОЛЕЙ БАЗ ДАННЫХ ЧЕРЕЗ REST API

Мочалов И.С.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

REST API повсеместно реализует метод обмена данными между пользователем и точками доступа сервера для получения контента на основе аутентификации. Отсутствие информационной безопасности возникает из-за включения необработанных полей из баз данных, таких как последовательные идентификаторы событий, заказов и т. д., для отображения пользователю, поскольку их можно использовать в качестве источника извлечения информации о показателях, связанных с операционной деятельностью компании. Основное внимание в работе уделяется подходам к анализу данных, применяемым к передаваемым полям в ответах API, чтобы продемонстрировать потенциальные уязвимости такой архитектуры, а также исследованию методов обеспечения контура безопасности информации.

Введение

Акцент на безопасность инфраструктуры в каналах связи между клиентом и сервером привел к значительным улучшениям в технологиях и правилах, реализованных для предотвращения генерации и изменения отправленных данных или блокировки возможности их получения со стороны злоумышленников. Основание проблемы даже при всех принятых

ограничениях неизменно остается тем же - сознательно выбранный для отправки контент может непреднамеренно также содержать уязвимости информационной системы.

Цель работы - показать уязвимость системы с полями с автоинкрементов, когда они выбираются для использования в качестве основного идентификатора во фронтенд-системах приложений и, таким образом, отправляются открыто через точки обмена информацией внутри API. Для достижения этой цели решается ряд задач исследования: анализ протоколов передачи данных клиент-сервер и взаимодействия систем управления базами данных (СУБД) и API, изучение наборов правил аутентификации для разделения доступа, определение автоинкремента в базах данных, построение шаблона для автоматизированного сбора идентификаторов, моделирование подходов к анализу данных для собранных значений, оценка возможных преимуществ для конкурентов от повторения рассмотренного исследования и обзор архитектурных решений для обеспечения информационной безопасности API.

Статья состоит из трех разделов, посвященных основам обработки информации, подходам к конкурентному анализу данных на основе парсинга и решениям для улучшения информационной безопасности REST API.

Обработка данных в коммуникации клиент-сервер

Общепринятый процесс обмена данных через API включает в себя операции аутентификации пользователей, ведущие к созданию пар bearer токенов по принципу OAuth методологии, для хранения на устройстве для будущей авторизации, реагированию на присланные указания к получению данных из подключенных баз и удалением или вставкой новых записей для поддержания актуальной информации о пользователе [1]. В примере мобильного приложения для краткосрочной аренды действиями, которые могут быть реализованы при помощи GET-запросов, которые осуществляют отправку данных с сервера пользователю, могут включать получение страницы с доступными для бронирования объектами.

Интеграция СУБД открывает возможность конвертировать сгенерированные действия пользователя через API в SQL-подобные запросы удобным и безопасным способом, скрывая команды и процедуры управления соединениями от конечного пользователя и фильтруя любые вредоносные запросы, которые могут быть получены. Например, при прикреплении личного токена аутентификации среди других заголовков к GET к пути точки доступа до истории бронирований с таблицей под названием «Bookings», хранящей информацию обо всех совершенных транзакциях в приложении, полученный запрос к backend затем внутри преобразуется в запрос, который может выглядеть следующим образом:

```
SELECT booking_id, listing_id  
FROM Bookings
```

*WHERE user_id = *user_id*;
 booking_id (Int64) - id транзакции бронирования
 listing_id (Int64) - id бронированного объекта
 user_id (Int64) - id пользователя, сделавшего бронирование*

Такой подход обеспечивает удобный метод взаимодействия между пользователем и сервером для представления необходимых данных в приложении с предотвращением доступа к не релевантной для имеющейся роли информации. Предварительная фильтрация входных параметров и внутренняя генерация скрипта обращения к БД становится дополнительным инструментом безопасности от атак типа SQL-injection [2].

Большинство СУБД предоставляют подход к организации хранения и взаимодействия с последовательностью записей в таблицах путем наличия механизма автоматического инкремента счетчика записей. Он основан на сохранении в памяти текущего значения счетчика *counter*, представляющего собой числовое значение индекса последней записи, записанной в таблицу, и присвоение значения *counter + 1* новой строке, появляющейся в таблице, и, наконец, сохранении обновленного значения в *counter* [3].

Описанный автоинкрементный идентификатор часто выступает в роли первичного ключа и включается в ответы на GET, с точки зрения бизнеса это позволяет упростить синхронизацию ссылок из разных частей продукта на одну и ту же запись в базе данных для любых пользователей, служб или устройств. Однако благодаря аналитическому подходу становится возможным использовать такие числовые идентификаторы в исследовании данных для получения ценных операционных показателей компании без необходимости технического вмешательства в систему.

Конкурентный анализ на основе автоинкрементных идентификаторов полей базы данных

Анализ метрик начинается с процесса сбора данных для получения значений выбранных измерений, привязанных к указанным временным меткам момента их сбора, что позволяет проводить дальнейшее исследование временных рядов. POST-запрос с данными транзакции, отправленный после того, как пользователь выполнил соответствующее действие в интерфейсе, возвращает json-объект, в котором среди других полей присутствует идентификатор транзакции, созданный в таблице внутренней базы данных для соответствующей записи, в следующей структуре: `{"datetime" : "YYYY-mm-dd HH:MM:SS", "transaction_id": 4190823068}`. Пример нескольких значений *transaction_id* в разные периоды времени (см. рис.1).

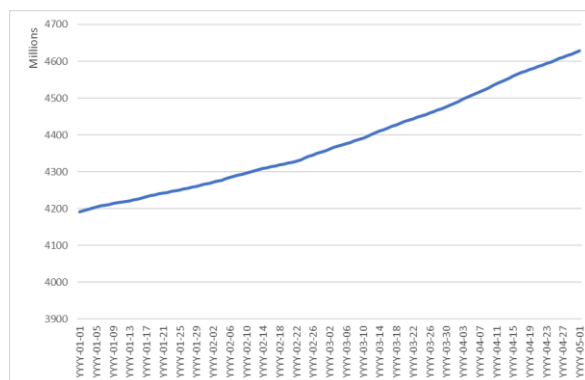


Рис.1. Динамика собранных значений инкрементного идентификатора

Абсолютные значения возвращаемых последовательных числовых полей не предоставляют полезной информации для описания соответствующей метрики, поскольку текущее значение, скорее всего, включает все итерации инкремента с момента самой первой транзакции, обработанной в выбранном backend решении, с отсутствием возможности отслеживать точный момент, когда эта система была инициализирована, вне некоторых исключений [4].

Скрипт автоматического сбора данных, предназначенный для генерации или просмотра выбранных идентификаторов с определенными интервалами, позволяет сравнивать, как изменялось значение в течение периодов измерения. Затем дельты рассчитываются как разница между текущими и предыдущими значениями, и представляются результаты сборов скриптом данных ежедневно с равными интервалами (см. таблицу 1).

Таблица 1. Значения дельт ID поля из БД

| datetime | transaction_id | id delta |
|------------|----------------|----------|
| YYYY-ММ-01 | 4190823068 | - |
| YYYY-ММ-02 | 4194606047 | 3446714 |
| YYYY-ММ-03 | 4198052761 | 3337648 |
| YYYY-ММ-04 | 4201390409 | 3225201 |
| YYYY-ММ-05 | 4204615610 | 2808664 |

Публичные компании, как правило, предоставляют ежеквартальную финансовую информацию о своей деятельности с включением в нее операционных количественных показателей. На примере компании по краткосрочной аренде в отчетности может быть представлено количество активных пользователей на платформе или добавленных новых объектов владельцами. Если компания публикует, что за тот же период, в который производился регулярный парсинг, было реализовано $0,75 \cdot x$ транзакций от x общего значения дельты, можно предположить, что четверть всех транзакций, записанных в базу данных, были недействительными или по каким-либо другим

причинам нерелевантными для выбранного исследования (если линейная регрессия среди других гипотез имеет наибольшее сходство с паттерном соотношения между точками данных из парсинга и сообщаемых в отчетности значений, с константой смещения, равной 0). Окончательный переход от исходных необработанных идентификаторов к измеримым временным рядам (см. рис.2).

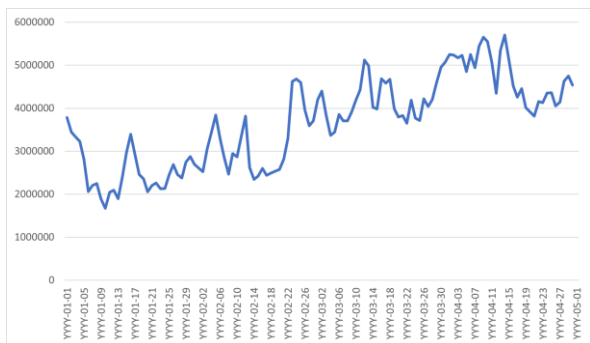


Рис.2. Временной ряд дельт значений ID

Сочетание парсинга информации из незащищенного API и проверки соответствующих отчетов приводит к появлению возможностей для принятия решений на основе данных со стороны конкурентов анализируемой компании. Метрики, построенные на основе нескольких ключевых параметров операционной деятельности компании, являются полезным источником сравнения рыночных тенденций в отрасли целиком, а также влияния выбранных рекламных кампаний или новых функций, реализованной на уровне всего бизнеса конкурентом или в целевых A/B-тестах, направленных на увеличение доли рынка в конкретных срезах пользователей или бизнеса. Более того, детальные периоды времени, в течение которых можно наблюдать изменения, а не только точечные значения раз в квартал из отчетностей, позволяя быстрее проверять гипотезы (выполняется проверка статистической значимости, чтобы доказать, что выбранная продолжительность периода наблюдения является репрезентативной для порядка общего прироста ID за отчетный квартал).

Архитектурные решения REST API для обеспечения безопасности информации

Основная уязвимость системы связана с возможностью измерения различий между значениями идентификаторов из-за их последовательной генерации, поэтому полезное решение с заменой полей автоинкремента на случайно сгенерированные является действенным. В такой архитектуре, если пользователь наблюдает несколько идентификаторов одного и того же поля базы данных, становится невозможным понять, сколько идентификаторов было сгенерировано между ними. Чтобы дополнительно обеспечить уникальность каждого созданного значения ID, можно использовать 128-битную стандартную разметку правил, используемую для хранения информации, называемую UUID, поскольку вероят-

ность ее дублирования достаточно близка к нулю, чтобы считать незначительной [5].

Переход от последовательного счетчика к случайному методу генерации идентификаторов может привести к неудобствам для ранее использовавшихся подходов к обработке данных на frontend, например, для сортировки возвращаемого списка объявлений, доступных для бронирования, от недавно добавленных к самым старым. Помимо многочисленных путей, открытых для решения таких проблем, можно рассмотреть вариант с искусственной необходимостью работать с сортировкой по идентификатору. Решением будет являться сохранение поля идентификаторов автоинкремента в базе данных одновременно со случайно созданными значениями и сортировка возвращаемых значений в backend с добавлением соответствующего условия сортировки в генерируемый SQL-подобный запрос к БД в ответ на присланные пользователем условия к API. Поле автоинкремента таким образом используется в тех же манипуляциях, но остается невидимым, поскольку никогда не отправляется вне СУБД.

Дополнительная безопасность информации может быть достигнута за счет архитектуры дерева точек доступа в REST API. Следование политике zero-tolerance в отношении сквозного доступа к данным, относящихся к пользователю А, со стороны пользователя В должно быть выбрано в качестве приоритета. Можно рассмотреть случай, когда путь доступа к API для GET запроса получения информации о транзакции структурирован как `./api/v2/transactions/{transaction_id}` и при возврате содержимого от БД для указанного `transaction_id` не применяются дополнительные правила фильтрации, исходя из аутентификации. При такой архитектуре существует возможность получить информацию о транзакции другого пользователя, зная ее идентификатор или путем его перебора. Дополнительно в случае, когда известно, что идентификатор сгенерирован с помощью правила автоинкремента, появляется возможность проверки ряда последовательных чисел на предмет наличия существующих транзакции с таким ID в системе, чтобы получить данные о них.

Внедрение дополнительного уровня проверок аутентификации при доступе к БД, хранящей соответствующие данные, позволяет отфильтровывать нарушения установленных логических правил. Например, пользователи должны иметь доступ к просмотру только своих собственных транзакций, что в SQL может быть реализовано изменением из состояния 1) в состояние 2) следующим образом:

- 1) `SELECT * FROM Transaction
WHERE transaction_id = *transaction_id*`
- 2) `SELECT * FROM Bookings
WHERE transaction_id = *transaction_id*
AND user_id = *user_id*`

Более глубокая переработка дерева точек доступа может включать в себя удаление всех URL-путей, в которых отправляются значения глобальных иден-

тификаторов, чтобы исключить попытки несанкционированного доступа к любым конфиденциальным данным, заменив их на локально сгенерированные для каждого пользователя. Дальнейшая обработка ссылки на глобальный идентификатор внутри backend и, таким образом позволяет упразднить отображение пользователю любых глобальных идентификаторов БД, как автоинкрементных, так и случайно-сгенерированных.

Анализ эффективности протокола обмена данными с модификацией архитектуры

Переход от автоинкрементов к другим видам счетчиков предполагает изменения в работе протокола обмена данными API и организации их хранения и генерации. В случае с автоинкрементом, являющимся типом данных integer или, как более распространенный и стабильный вариант для избежания переполнения, bigint, СУБД будет выделять 4 или 8 байт соответственно для хранения каждого значения. Переход к 128-битовому идентификатору UUID приведет к росту ассоциируемого под идентификатор пространства в 2 раза (относительно bigint) до 16 байт на запись. Большой рост занимаемой памяти произойдет в случае архитектуры с гибридным одновременным использованием автоинкрементов исключительно для внутренних расчетов и псевдослучайных идентификаторов для отображения вне базы данных. Этот подход добавит новое поле с увеличением памяти для идентификаторов таблицы в 2.5 раза, но позволит решить уязвимость информационной системы и поддержать legacy-логику предыдущей организации СУБД за счет обратной связи с ранее созданными автоинкрементами через сопоставление каждому нового идентификатора, который эксклюзивно в дальнейшем будет передаваться через API.

Кроме изменений в памяти, выделяемой на хранение идентификаторов, при переходе от автоинкрементов также будет затронута производительность генерации нового ID. Большинство СУБД реализует метод создания автоинкрементного значения на системном уровне и основывает логику на получении следующего числа в последовательности с хранением крайнего сгенерированного в памяти для предотвращения коллизий в случае удаления записей из БД. Генерация UUID предполагает более сложные математические операции, для псевдослучайного идентификатора версии 4, например, основывается на генерации двух 32-бит чисел для составления при помощи их объединения seed-значения функции генерации случайной последовательности с последующей операциями деления на степень 2 и смещения лишних значений битов полученной бинарной величины. Дополнительные вычислительные ресурсы будут использоваться при операциях преобразования значения из оптимального для хранения внутри БД к текстовому для передачи в API и обратно, а также при реализации автоматического создания значения при помощи триггера, активируемого при записи в таблицу и записывающего сгенерированное значение поверх переданного пустого, вместо нативной встав-

ки значений напрямую. В рамках оценки производительности предложенной модификации протокола было измерено среднее время создания и записи (транзакции) одного идентификатора при симуляции генерацией СУБД большого количества автоинкрементов и UUID. Результат на среднем и большом объеме транзакций отражает разницу в 3.5-7 раз во времени реализации за единицу в пользу автоинкремента (см. таблицу 2).

Таблица 2.

Бенчмарк скорости генерации разных идентификаторов

| Объем данных \ Идентификатор | Автоинкремент | UUIDv4 |
|---|----------------|----------------|
| Малый (в рамках одного буфера памяти) | 0.0011 секунды | 0.0012 секунды |
| Средний (больше одного буфера, но в рамках RAM) | 0.0012 секунды | 0.0038 секунды |
| Большой (больше RAM) | 0.0012 секунды | 0.0079 секунды |

Передача данных по протоколу API модификация с внедрением нового идентификатора влечет рост объема трафика для большинства случаев, так как при преобразовании в json-формат UUID занимает 36 символов вместе с разделителями в сравнении с не более чем 12-15 для поля автоинкремента при оценке среднего количества записей в таблицах. Таким образом, совокупный анализ модификации протокола по памяти, объему передаваемых данных и эффективности генерации отображает линейное увеличение потребляемых ресурсов для случайного идентификатора относительно автоинкремента.

За счет перехода от автоинкремента к UUID в дополнение к устранению уязвимостей системы появляются дополнительные возможности: децентрализация генерации идентификатора позволяет проще мигрировать таблицы между СУБД и реализовывать шардинг для их распределенного хранения, а также переносить нагрузку генерации на сторону пользователя при необходимости.

Заключение

Результаты исследования приводят к подтверждению передачи автоинкрементных полей БД как небезопасного метода обмена данными, поскольку его можно использовать для конкурентного анализа и получению статистики об операционной деятельности рассматриваемой компании. Извлечение такой информации имеет негативные долгосрочные последствия в случае ее применения для разработки стратегии, направленной на получение доли рынка со стороны конкурентов. Прикладная новизна данной работы состоит в разработке модификации протокола для повышения уровня информационной безопасности архитектуры REST API, включающей замену открытых автоинкрементных идентификаторов

на случайно генерируемые со связью только в бэкенде, а также реструктуризацию дерева точек доступа для предотвращения несанкционированного доступа и извлечения данных путем перебора и массового парсинга.

Дальнейшая работа, основанная на исследовании в этой статье, сосредоточена на более дополнительных методах эксплуатации уязвимостями архитектур API, таких как брутфорс небезопасно сгенерированных UUID, а также на способах улучшения инфраструктурой стабильности и защищенности информации путем фильтрации трафика, ассоциирующегося с автоматическим парсингом.

Список литературы

1. Bhat, P. K., & Kansal, R. Development of RESTful Web API using Token-based OAuth 2.0 Authorisation // International Journal of Engineering Research. International Journal of Engineering Research & Technology (IJERT) // Vol. 9 // Issue 10 - 2020.
2. Nasereddin, M., ALKhamaiseh, A., Qasaimeh, M., & Al-Qassas, R. A systematic review of detection and prevention techniques of SQL injection attacks // Information Security Journal: A Global Perspective // 32(4) // 252-265. - 2023.
3. Huang, J., Guo, J., Zhang, Z., Qian, W., & Zhou, A. Efficient Auto-Increment Keys Generation for Distributed Log-Structured Storage Systems // In Web Information Systems Engineering–WISE 2018: 19th International Conference, Dubai, United Arab Emirates, November 12-15, 2018, Proceedings, Part II 19 (pp. 225-239) // Springer International Publishing - 2018.
4. Dudjak, M., & Martinović, G. An API-first methodology for designing a microservice-based Backend as a Service platform // Information Technology and Control //49(2) // 206-223 - 2020.
5. Leach, P., Mealling, M., & Salz, R.. A universally unique identifier (uuid) urn namespace (No. rfc4122) - 2005.

СРАВНЕНИЕ ПРОТОКОЛОВ ПЕРЕДАЧИ ДАННЫХ WEBSOCKET И HTTP/2.0

Вербенко С.С.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В данной работе представлено сравнение производительности протоколов WebSocket и HTTP/2.0 в контексте передачи данных в реальном времени. Исследование оценивает эффективность работы протоколов при передаче данных разного характера и выявляет преимущества и недостатки каждого из них в различных сценариях использования. Полученные результаты могут быть полезны для разработчиков и инженеров при выборе оптимального протокола для конкретных веб-приложений.

Введение

С развитием веб-приложений и увеличением потребностей в передаче данных в реальном времени становится критически важным выбор оптимального протокола для обеспечения высокой производительности и эффективной коммуникации между клиентом и сервером. Целью данного исследования является сравнение производительности двух ключевых протоколов: WebSocket [1] и HTTP/2.0 [2], в контексте их использования для передачи потоковой информации и взаимодействия с веб-приложениями.

В данной работе оценивалось время отклика серверов, реализующих передачу данных по протоколам WebSocket и HTTP/2.0.

Обзор технологий

HTTP/2.0 - это обновленная версия протокола HTTP, которая призвана увеличить производительность веб-приложений путём оптимизации передачи данных между клиентом и сервером. Этот протокол базируется на предыдущей версии, HTTP/1.1, однако внедряет ряд новых функций и улучшений, направленных на уменьшение задержек, снижение нагрузки на сеть и повышение эффективности передачи ресурсов:

- Мультиплексированный запрос. Не существует ограничений на количество одновременных запросов, которые могут быть отправлены по одному подключению

- Приоритетные запросы. Клиенты могут запрашивать определённые ресурсы, которые должны быть доставлены в первую очередь, что позволяет избежать перегрузки сетевого тракта невостребованными ресурсами при наличии высокоприоритетного запроса.

- Сжатый заголовок. В HTTP/2.0 используется сжатие заголовков hpack [3].

- Server pushed streams – этот механизм позволяет серверу передавать данные клиенту, которые не были запрошены, но, которые ему будут необходимы в дальнейшей работе.

Протокол передачи данных WebSocket обеспечивает двустороннее бинарное взаимодействие между клиентом и сервером через постоянное соединение. Он разработан для преодоления ограничений, характерных для обычных HTTP-запросов, таких как избыточные заголовки и ограниченная возможность работы в режиме реального времени и передачи потоковых данных. Размеры заголовка у данного протокола составляют от 2 до 16 байт, что значительно меньше по сравнению со среднестатистическим заголовком в 550 байт у HTTP/2.0 [4]. Основным отличием протокола WebSocket от протокола HTTP/2.0 является обеспечение постоянного полнодуплексного соединения между клиентом и сервером, что предоставляет возможность обеим сторонам отправлять сообщения друг другу в режиме реального времени, в то время как HTTP/2.0 оптимизирует запрос-ответную модель для эффективной передачи ресурсов веб-приложений.

Реализация серверов

Для проведения эксперимента для каждой из двух технологий был реализован сервер на языке Python с использованием библиотеки Quart [5]. Данная библиотека предоставляет возможность для реализации сервера, который способен принимать и отправлять данные как по протоколу HTTP/2.0, так и по протоколу WebSocket. Для отправки большого количества одновременных запросов использовался инструмент для проведения нагрузочного тестирования Apache Jmeter [6].

И сервер на Python, и Apache Jmeter при проведении экспериментов были запущены на одной физической машине. Таким образом накладных расходов на физическую передачу данных нет (рис.1).

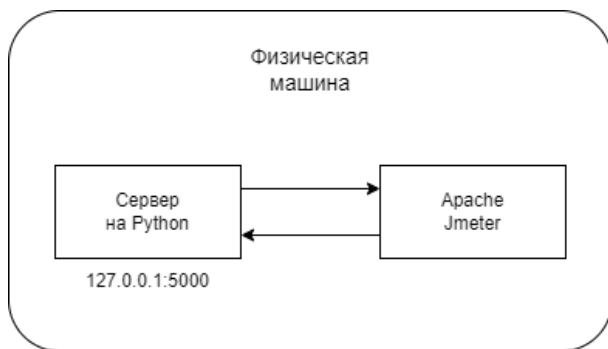


Рис.1. Архитектура тестового стенда

Технические характеристики машины, на которой проводился эксперимент:

- Оперативная память 16 ГБ
- Процессор AMD Ryzen 5 5600U 2.30 GHz
- Операционная система Windows 11

Первый сценарий

Для каждого из протоколов было запланировано два сценария для оценки производительности в различных условиях нагрузки. Первый сценарий включал отправку большого количества запросов от множества параллельных пользователей. В этом сценарии каждый клиент отправлял на сервер запрос, на который сервер отвечал полезной информацией (без учёта заголовков) размером 28 байт.

Для каждого протокола было выполнено семь измерений времени отклика. В каждом измерении создавалось разное количество параллельных пользователей. Каждый пользователь отправлял по 100 запросов, что позволяло оценить поведение протоколов при различных уровнях нагрузки и выявить их производительность в условиях высокой нагрузки. Результаты измерений времени отклика, проведённых по первому сценарию приведены на рис.2.

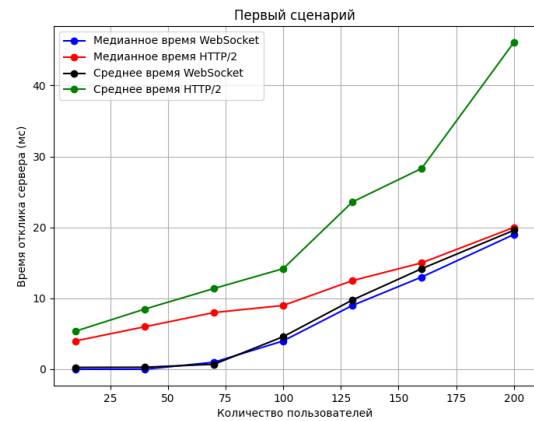


Рис.2. Время отклика для разного количества пользователей при малых запросах

В сформированных программой Apache Jmeter отчётах можно было найти размер отправленной сервером информации. По протоколу WebSocket отправлялось 30 байт данных в каждом ответе от сервера. В случае с протоколом HTTP/2.0, ответ содержал 267 байт. Так как полезной информации в ответе содержалось 28 байт, то по протоколу HTTP/2.0 передаётся 239 байт избыточной информации, в то время как по протоколу WebSocket 2 байта.

Второй сценарий

Второй сценарий проведения измерений заключался в отправке большого количества информации в одном ответе от сервера. В данном случае также было произведено пять замеров для каждого протокола с разным количеством параллельных пользователей. В каждом из трёх экспериментов от одного пользователя поступало 10 последовательных запросов. На запрос от клиента сервер отправлял JSON размером 1.9 МБ. Размер ответа обусловлен ограничением буфера Apache Jmeter в чуть более, чем 2 МБ. Время отклика у сервера, работающего по протоколу WebSocket, оказалось меньше практически во всех случаях (рис.3).

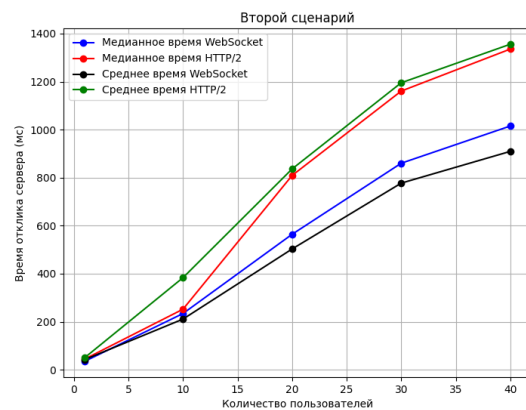


Рис.3. Время отклика для разного количества пользователей при больших запросах

Безопасность

Помимо производительности, важным аспектом при выборе между протоколами WebSocket и HTTP/2.0 является безопасность. Оба протокола предоставляют различные механизмы для обеспечения безопасности соединения и передаваемых данных.

По умолчанию WebSocket не предоставляет механизмов шифрования данных или аутентификации. Для обеспечения безопасности соединения WebSocket может быть использовано соединение по протоколу WebSocket Secure (wss://), которое использует шифрование TLS/SSL для защиты данных. Протокол не предписывает никакого конкретного способа, с помощью которого серверы могут аутентифицировать клиентов во время подключения. Для аутентификации пользователей и обеспечения безопасности данных в WebSocket может использоваться дополнительная авторизация на уровне приложения.

HTTP/2.0 поддерживает использование защищённого протокола передачи данных - HTTPS (HTTP over TLS), который обеспечивает шифрование данных между клиентом и сервером с помощью протокола TLS. Это защищает конфиденциальность и целостность передаваемой информации, предотвращая прослушивание и подмену данных третьими лицами. Дополнительные механизмы безопасности, такие как межсайтовая защита от подделки запросов (CSRF) и защита от атаки инъекцией (XSS, SQL Injection), могут быть реализованы на уровне приложения.

Заключение

В данной статье было проведено сравнение производительности протоколов WebSocket и HTTP/2.0 в контексте передачи данных в реальном времени. Исследование позволило оценить эффективность работы каждого из протоколов при передаче данных разного характера и выявить их преимущества и недостатки в различных сценариях использования.

В случае, когда множество пользователей передаёт информацию небольшими запросами, WebSocket обычно демонстрирует более эффективное использование трафика и более быструю обработку по сравнению с HTTP/2.0. Однако, при увеличении количества параллельных пользователей, разница между протоколами становится практически незаметной, так как графики медианного времени отклика для двух протоколов почти пересеклись на уровне 200 одновременных пользователей.

Когда передаётся большой объем данных в одном запросе, WebSocket все ещё имеет преимущество благодаря сокращению накладных расходов на трафик и незначительному выигрышу в скорости обработки.

Безопасность рассматриваемых протоколов обеспечивается использованием их вместе с протоколом TLS. Это обеспечивает конфиденциальность и целостность информации, а также проверку подлинности источников данных. Тем не менее внедрение рассмотренных протоколов должно сопровождаться реализацией механизмов авторизации и аутентификации для проверки подлинности пользователей и

ограничения доступа к ресурсам, а также экранированием данных и использованием CSRF-токенов.

Таким образом, выбор между WebSocket и HTTP/2.0 зависит от конкретных требований и сценариев использования. Несомненным преимуществом протокола WebSocket является двунаправленное соединение, которое позволяет серверу отправлять информацию без необходимости ожидать запрос от клиента.

Список литературы

1. Fette I., Melnikov A. The websocket protocol. – 2011. – №. rfc6455.
2. Belshe M., Peon R., Thomson M. Hypertext transfer protocol version 2 (HTTP/2). – 2015. – №. rfc7540.
3. Peon R., Ruellan H. HPACK: Header compression for HTTP/2. – 2015. – №. rfc7541.
4. de Souza Soares E. F. et al. Evaluation of Server Push Technologies for Scalable Client-Server Communication // 2018 IEEE Symposium on Service-Oriented System Engineering (SOSE). – IEEE, 2018. – С. 1-10.
5. Quart 0.19.4 documentation [Электронный ресурс]. URL: <https://quart.palletsprojects.com/en/latest/index.html> (дата обращения: 29.01.2024).
6. User's Manual // Apache Jmeter [Электронный ресурс]. URL: <https://jmeter.apache.org/usermanual/index.html> (дата обращения: 29.01.2024).

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ НЕЙРОННЫХ СЕТЕЙ К ПОСТРОЕНИЮ УНИВЕРСАЛЬНЫХ АТАК НА ХЕШ ФУНКЦИИ

Чибрин М.Д.

*Национальный исследовательский университет
«Высшая школа экономики»,
кафедра информационной
безопасности киберфизических систем
МИЭМ НИУ ВШЭ*

Аннотация

В работе предлагается использование нейронных сетей для анализа возможности применения их в атаках на хеш-алгоритмы. Данное исследование проводится в рамках разработки модели на основе нейронных сетей, способной отличить случайную от псевдослучайной последовательности, полученной на основе поточных алгоритмов хеширования

Введение

Важность значения псевдослучайных последовательностей проявляется при решении различных задач, связанных с исследованием информационных моделей, анализом работы программного и аппаратного обеспечения, реализацией надежной передачи, обработки и хранения информации.

Анализ псевдослучайной последовательности в алгоритмах хеширования имеет огромное значение

при решении задач в области криптологии. Большинство хеш-функций основано на поточных итеративных функциях, преобразующих последовательность из «К» входных бит в последовательность ограниченной длины бит «N», где «N» - разрядность хеш-функции. Выходная последовательность бит зависит от всего входного потока данных и начального значения.

В данной работе рассматривается возможность применения моделей нейронных сетей, способных различать соседние раунды внутри алгоритма хеширования. Основной задачей исследования является нахождение максимального числа итераций, при котором модель еще может отличить соседние раунды.

Атака-различитель (англ. Distinguishing attack)

Для распознавания псевдослучайных последовательностей, полученных после различного количества раундов, используются решающие правила на основе различителей. Определение максимального количества раундов, при котором такие правила могут отличить разные раунды друг от друга, является основной задачей атаки-различителя. Анализ, применяемый в «атаке-различителе», схож с подходом в «Дифференциальной атаке» (англ. Differential Attacks) [1,2], в которой исследуется зависимость выходной последовательности от подаваемых на вход символов. Однако, основная разница между этими подходами в том, что при исследовании алгоритма с помощью дифференциальной атаки наблюдателю не важно, что происходит внутри алгоритма, после очередного раунда. Выбранный в исследовании способ анализа криптоалгоритмов имеет более углубленный подход.

Одна из основных проблем криптографии – сохранение баланса между скоростью алгоритма и его криптографической стойкостью. «Атака-различитель» позволяет исследовать как промышленные, так и упрощенные версии алгоритмов, чтобы определить минимальное количество раундов преобразования открытого текста в последовательность символов, ничем не отличную от случайного набора символов [3].

Большинство исследований о возможности применения атак различителей было проведено над алгоритмами шифрования, таких как Simon [4], AES [5], RC4 [6], но было и исследование в для алгоритма хеширования Speck32/64 [7]. Данное исследование позволит расширить область применения данной атаки на поточные алгоритмы хеширования sha-1 и sha-256.

Предобработка данных

Как уже было сказано ранее, в работе исследуются два алгоритма хеширования: sha-1 и sha-256. Алгоритмы SHA-1 (Secure Hash Algorithm 1) и SHA-256 (Secure Hash Algorithm 256) являются криптографическими хеш-функциями, которые используются для вычисления уникального хеш-значения для входных данных.

В ходе решения задачи по предобработке входных данных был предложен следующий алгоритм по

преобразованию шестнадцатеричного текста в RGB изображение:

1. Шестнадцатеричная последовательность преобразуется в массив из чисел по принципу 2 символа «хеш-текста» = 1 числу.
2. Длина числового массива берется за ширину и высоту изображения.
3. Каждый пиксель изображения заполняется тремя последовательными числами из массива.
4. Так как чисел намного меньше пикселей, то данные используются повторно, пока строка не будет заполнена.
5. Строки повторяются пока изображение полностью не заполнится.
6. Размер изображения изменяется до 299x299 пикселей.

Для примера рассмотрим первые 32 итерации внутри алгоритма sha-1 (рис.1) и sha-256 (рис.2) визуализированные с помощью предложенного алгоритма.

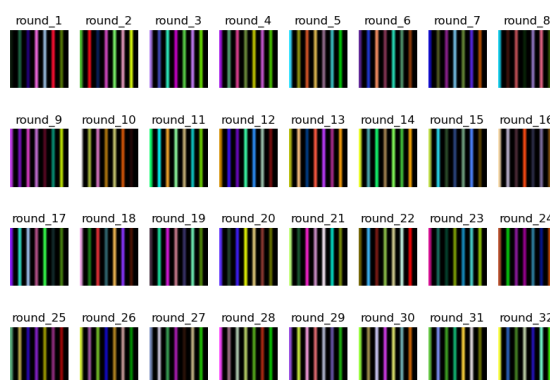


Рис.1. Первые 32 итерации алгоритма sha-1, визуализированные с помощью обработчика

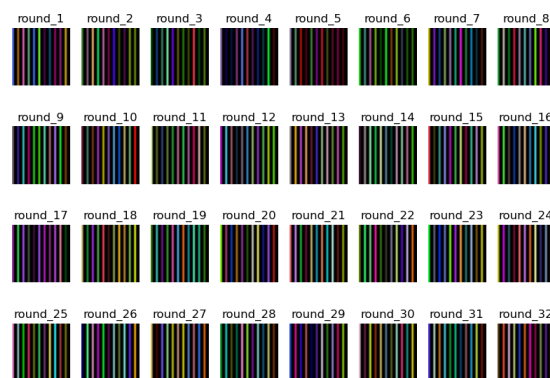


Рис.2. Первые 32 итерации алгоритма sha-256, визуализированные с помощью обработчика

Чтобы наглядно обосновать проводимое исследование, стоит сравнить выборку из сгенерированных изображений первого раунда и второго (рис.3). Даже невооруженным взглядом видно, что эти раунды не похожи друг на друга. В первой и во второй строках содержатся попарно две разные визуализированные последовательности двух алгоритмов sha-1 и sha-256

соответственно. Изображения схожи между собой по раундам в каждом алгоритме, но в тоже время присутствуют и различия. Чем выше будет раунд, тем различия будут заметнее. Это нам подтвердит и сама модель, о которой в следующем блоке.

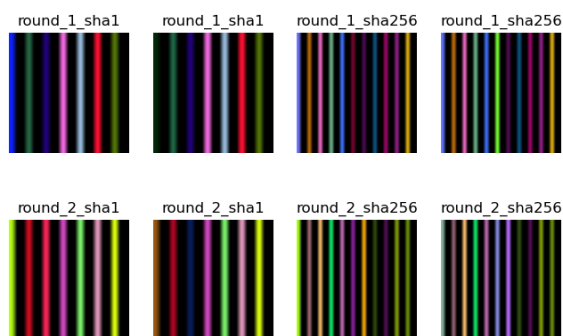


Рис.3. Наглядное различие в раундах

Анализ данных и получение результатов

Нейронной сетью, на основе которой разработана модель, способная отличить случайную последовательность N-го раунда от случайной последовательности N+1-го раунда, была выбрана Inception_V3 (рис.4).

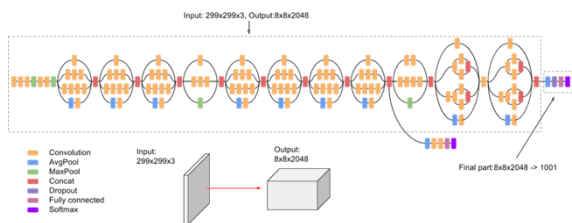


Рис.4. Архитектура сверточной нейронной сети Inception_V3

На основе данной нейронной сети уже были проведены множество исследований в области криптографии [8,9], но самое важное для нас то, что эта нейросеть уже применялась в схожих условиях, но в качестве объекта для анализа выступали блочные итеративные алгоритмы шифрования [4,7,10]. Исходя из проведенных ранее исследований, настроим выбранную модель на формат подготовленных ранее данных.

Для получения приведенных ниже результатов были использованы следующие параметры для обучения модели:

1. Размер выборки (англ. Batch size) – 25
2. Количество процессов (англ. Workers) – 5
3. Количество эпох (англ. Epochs) – 10 (данное значение выбрано как максимальное, если точность определения резко падает, то обучение останавливается и берется последнее значение)

В ходе реализации модели были получены следующие результаты (рис.5 и рис.6)

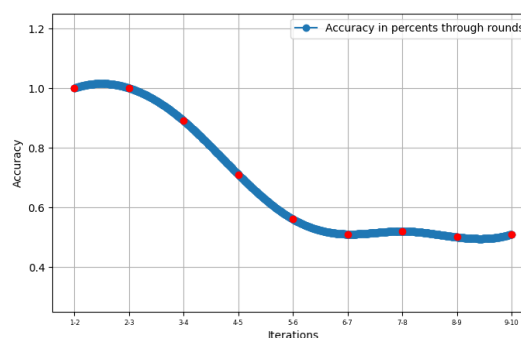


Рис.5. Показатели точности определения раундов модели при исследовании итераций внутри алгоритма sha-1

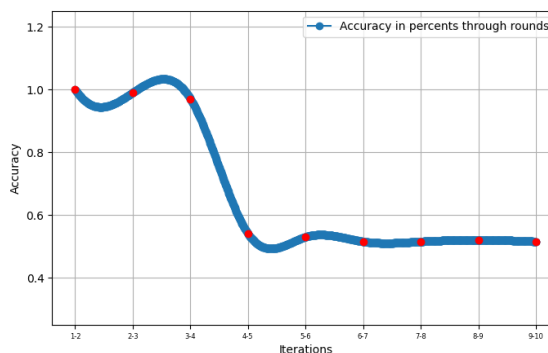


Рис.6. Показатели точности определения раундов модели при исследовании итераций внутри алгоритма sha-256

По итогу обучения модели и проведения статистических тестов были получены следующие результаты:

1. Уже после 5-й итерации в алгоритме sha-1 модель перестает различать раунды между собой с высокой точностью, а следовательно максимальный искомый раунд для данного алгоритма равен 5.

2. В свою очередь искомый раунд для алгоритма sha-256 равен, что говорит о его большей стойкости к проводимой атаке.

Заключение

В результате проведенных исследований, была предложена модель на основе сверточной нейронной сети Inception_V3, которая способна различить результаты хеширования между двумя последовательными итерациями внутри алгоритма.

Список литературы

1. B. Preneel, B. Van Rompay, L. Granboulan, G. Martinet, S. Murphy, R. Shipsey, J. White, M. Dichtl, P. Serf, M. Schafheutle, E. Biham, O. Dunkelman, V. Furman, M. Ciet, J-J. Quisquater, F. Sica, L. Knudsen, H. Raddum / Security Evaluation of NESSIE First Phase 2001 // IST-1999-12324 : электронный журнал. – URL: https://www.researchgate.net/profile/Jean-Jacques-Quisquater/publication/2545150_NESSIE_Phase_I_Selection_of_primitives/links/02bfe510019956bb77000000/NESSIE-Phase-I-Selection-of-primitives.pdf (дата обращения 25.11.2023)
2. Knudsen L., Meier W. / Correlations in RC6 with a Reduced Number of Rounds // Fast Software Encryption – 2000. – p. 94–108 – URL: <https://link.springer.com/chap->

$$p(a_0 a_1 \dots a_s) = P(x_s = a_0 | x_{s-1} = a_1, \dots, x_0 = a_s)$$

Введем следующие обозначения:

- A – массив вероятностей для состояний глубины S : $p(a_0 \dots a_s)$;
- B – массив вероятностей для состояний глубины $S+1$: $p(a_0 \dots a_{s+1})$;
- C – массив вероятностей для состояний глубины $S+1$: $p(a_0 \dots a_{s+1})$.

Для вычисления вероятностей необходимо выполнить подсчет частот состояний цепи. Предварительно задается некоторое значение n , которое больше, чем предполагаемая глубина. Подсчет состояний производится, начиная с глубины 1 и до глубины n .

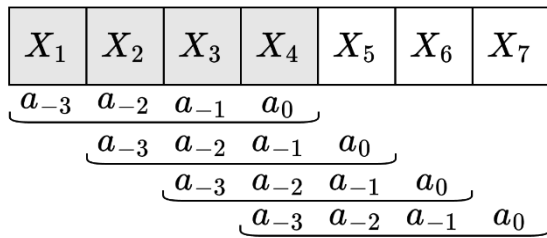


Рис. 1. Схема подсчета частот состояний для глубины 3

После подсчета частот состояний для каждой глубины производится вычисление условных вероятностей. Берется значение частоты состояния $f(a_0 \dots a_s)$ для глубины S , и значение частоты $f(a_0 \dots a_{s+1})$ для глубины $S+1$. Затем вычисляются вероятности состояния $(a_0 \dots a_{s+1})$ при условии предыдущих $a_0 \dots a_s$ элементов.

Для сравнения вероятностей состояний воспользуемся тремя методами: метрикой χ^2 , дивергенцией Кульбака-Лейблера, методом оценки доверительного интервала вероятностей. Решение выносится по результатам мажоритарного голосования трех методов: если любые 2 метода дают одинаковый результат, то выносится решение о глубине. Иначе, глубина не определена.

Для каждого метода будут приведены результаты определения глубины для выборки с глубиной 7 объемом 10^8 бит.

Метрика Хи-квадрат

Метрикой χ^2 называют функционал:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

где O – массив эмпирических значений, E – массив теоретических значений, n – количество состояний (значений) [1].

Значение расстояния χ^2 дает меру непохожести эмпирических значений на теоретические. Следовательно, чем меньше значение, тем более похожи (близки) распределения, задаваемые массивами O и E .

В данной работе вычисляется расстояние хи-квадрат $\chi^2(A, B)$, $\chi^2(A, C)$, и $\chi^2(B, C)$ для глубин от 1 до n . Затем в каждом из трех полученных массивов расстояний производится поиск минимального значения. Критерий принятия решения формулируется следующим образом: если большинство из трех индексов минимальных значений совпадает, то делается вывод, что глубина цепи равняется этому индексу. В противном случае глубина считается не определенной. Ниже приведены результаты эксперимента по определению глубины. Наименьшее значение статистик соответствует глубине 7.

Таблица 1. Значения метрики Хи-квадрат

| Глубина | $\chi^2(A, B)$ | $\chi^2(A, C)$ | $\chi^2(B, C)$ |
|---------|-----------------|-----------------|-----------------|
| 1 | 5.45E-04 | 5.97E-04 | 2.28E-03 |
| 2 | 4.29E-02 | 4.34E-02 | 1.72E-01 |
| 3 | 1.93E-01 | 2.60E-01 | 8.51E-01 |
| 4 | 4.51E-01 | 9.35E-01 | 3.25E+00 |
| 5 | 6.43E+00 | 4.75E+00 | 2.58E+01 |
| 6 | 2.09E+01 | 3.16E+01 | 5.48E+01 |
| 7 | 1.81E-04 | 3.77E-04 | 7.28E-04 |
| 8 | 1.01E-03 | 5.89E-03 | 8.63E-03 |
| 9 | 4.53E-02 | 3.74E-02 | 1.01E-01 |
| 10 | 2.07E-01 | 2.57E-01 | 6.82E-01 |

Дивергенция Кульбака-Лейблера

Дивергенцией (расхождением) Кульбака-Лейблера распределения P по отношению к распределению Q называется функционал:

$$D_{KL}(P||Q) = \sum_{i=1}^n p(x_i) \log \frac{p(x_i)}{q(x_i)}$$

где P и Q – дискретные вероятностные распределения, n – количество исходов [2].

Дивергенция характеризует то, насколько одна вероятностная схема плохо описывает другую вероятностную схему (чем значение больше, тем хуже).

В данной работе вычисляются дивергенции $D_{KL}(A||B)$ и $D_{KL}(B||A)$, $D_{KL}(A||C)$ и $D_{KL}(C||A)$, $D_{KL}(B||C)$ и $D_{KL}(C||B)$ для глубин от 1 до n . (В силу несимметричности расхождения вычисляется 6 статистик). Затем в каждом из шести полученных массивов производится поиск минимального значения. Критерий принятия решения: если большинство из шести индексов минимальных значений совпадает, то делается вывод, что глубина цепи равняется этому индексу. В противном случае глубина не определяется. В таблице 2 и 3 приведены результаты значений дивергенций, которые показывают, что проверяемая последовательность имеет глубину 7.

Таблица 2.
Значения дивергенции Кульбака-Лейблера

| Глубина | $D_{KL}(A B)$ | $D_{KL}(B A)$ | $D_{KL}(A C)$ |
|----------|-----------------|-----------------|-----------------|
| 1 | 2.72E-04 | 2.72E-04 | 2.99E-04 |
| 2 | 1.05E-02 | 1.04E-02 | 1.07E-02 |
| 3 | 2.16E-02 | 2.00E-02 | 3.01E-02 |
| 4 | 2.59E-02 | 2.48E-02 | 4.21E-02 |
| 5 | 1.37E-01 | 1.16E-01 | 9.75E-02 |
| 6 | 1.36E-01 | 9.58E-02 | 1.94E-01 |
| 7 | 1.42E-06 | 1.42E-06 | 2.95E-06 |
| 8 | 3.97E-06 | 3.98E-06 | 2.28E-05 |
| 9 | 8.67E-05 | 8.52E-05 | 6.64E-05 |
| 10 | 2.07E-04 | 2.14E-04 | 2.40E-04 |

Таблица 3.
Значения дивергенции Кульбака-Лейблера

| Глубина | $D_{KL}(C A)$ | $D_{KL}(B C)$ | $D_{KL}(C B)$ |
|----------|-----------------|-----------------|-----------------|
| 1 | 2.99E-04 | 1.14E-03 | 1.14E-03 |
| 2 | 1.05E-02 | 4.21E-02 | 4.21E-02 |
| 3 | 2.89E-02 | 9.78E-02 | 1.02E-01 |
| 4 | 3.42E-02 | 1.33E-01 | 1.16E-01 |
| 5 | 7.73E-02 | 4.14E-01 | 3.74E-01 |
| 6 | 1.52E-01 | 4.21E-01 | 4.35E-01 |
| 7 | 2.96E-06 | 5.69E-06 | 5.69E-06 |
| 8 | 2.27E-05 | 3.35E-05 | 3.33E-05 |
| 9 | 6.13E-05 | 1.92E-04 | 1.88E-04 |
| 10 | 2.34E-04 | 6.26E-04 | 6.05E-04 |

Определение доверительного интервала

Введем случайную величину $\xi_{1,i}$, которая будет равна 1, если в последовательности после состояния $(a_0 \dots a_s)$ следует единица. И случайную величину $\xi_{0,i}$, которая будет равна 1, если в последовательности после состояния $(a_0 \dots a_s)$ следует ноль. Обозначим эти два случая следующим обозначением: $\xi_{*,i}$, где * — это «ноль» или «единица».

Определим допустимое отклонение β двух вероятностей $p(a_0 \dots a_s 0)$ и $p(a_0 \dots a_s 1)$ от $p(a_0 \dots a_s)$ с заданным уровнем точности γ . Под отклонением будем понимать абсолютную разность между вероятностями. То есть получим границы промежутков $[p(a_0 \dots a_s 0) - \beta, p(a_0 \dots a_s 0) + \beta]$ и $[p(a_0 \dots a_s 1) - \beta, p(a_0 \dots a_s 1) + \beta]$.

$$P\left(p(a_0 \dots a_s *) - \beta \leq \frac{\sum_i \xi_{*,i}}{n} \leq p(a_0 \dots a_s *) + \beta\right)$$

Опираясь на теорему Муавра-Лапласа о сходимости биномиального распределения к нормальному [3] получим:

$$\beta \geq \Phi^{-1}\left(\frac{\gamma + 1}{2}\right) \sqrt{\frac{p(a_0 \dots a_s *) (1 - p(a_0 \dots a_s *))}{f(a_0 \dots a_s *)}}$$

Граница β показывает допустимое отклонение вероятностей состояний $p(a_0 \dots a_s)$ при глубине S от $p(a_0 \dots a_s *)$ при глубине $S + 1$.

Установим уровень точности γ равным 99%, который соответствует вероятности того, что истинное значение попадает в доверительный интервал. Для всех состояний проверяемой глубины считается количество состояний, для которых выполняется следующее неравенство:

$$|p(a_0 \dots a_s) - p(a_0 \dots a_s *)| \leq \beta.$$

Критерий принятия решения: определим, что если доля верных неравенств больше 90%, то принимается решение, что глубина последовательности равна проверяемой. Из таблицы ниже видно, что последовательность имеет глубину 7 (первая проверяемая глубина, которая соответствует критерию).

Таблица 4.

Доля попадания частот в доверительный интервал

| Глубина | Доля попадания в интервал |
|----------|---------------------------|
| 1 | 0.000 |
| 2 | 0.000 |
| 3 | 0.000 |
| 4 | 0.000 |
| 5 | 0.000 |
| 6 | 0.023 |
| 7 | 0.941 |
| 8 | 0.910 |
| 9 | 0.927 |
| 10 | 0.925 |

Преобразование последовательности

Для преобразования марковской последовательности в статистически неотличимую от равновероятной в работе применяется арифметическое кодирование [4], которое позволяет уменьшить информационную избыточность. Под последовательностью, статистически неотличимой от равновероятной, будем подразумевать последовательность, обладающую теми же статистическими свойствами, что и равновероятная. Результат преобразования проверяется с помощью набора статистических тестов NIST [5]. Преобразованная последовательность проходит набор статистических тестов.

Заключение

В ходе выполнения работы рассмотрена и решена задача по проведению численных экспериментов по проверке статистических гипотез марковости. Результаты численных экспериментов показали, что приведенные в работе методы позволяют определить глубину цепи Маркова. Кроме того, выполнена задача по преобразованию цепи Маркова в последовательность, статистически неотличимую от равновероятной. Марковская последовательность, сжатая с помощью примененного в работе алгоритма арифметического кодирования, успешно проходит набор статистических тестов.

Список литературы

1. Dodge Y. The concise encyclopedia of statistics. — 2010-е изд. — New York, NY : Springer, 2008. — 622 с.
2. Thomas M. Cover, Thomas J. A. Elements of Information Theory. — 2-е изд. — Nashville, TN : John Wiley & Sons, 2006. — 776 с.
3. Ивченко Г. И., Медведев Д. И. Введение в математическую статистику: учебник — М.:ЛКИ. — 2009. — 600 с.
4. Sayood K. Introduction to Data Compression. — 3-е изд. — Oxford, England : Morgan Kaufmann, — 2005. — 704 с.
5. A statistical test suite for random and pseudorandom number generators for cryptographic applications : tech. отч. / L. E. Bassham III [и др.]. — Gaithersburg, MD, 2010.

СРАВНЕНИЕ МОДЕЛЕЙ ИНФОРМАЦИОННОГО ВОЗДЕЙСТВИЯ В СЕТИ TELEGRAM-КАНАЛОВ

Попов В.А.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ*

Аннотация

В работе рассматриваются модели построения графа взаимодействующих объектов для сети информационного воздействия Telegram-каналов, в основе которых лежит функция подсчета весов ребер. Для 12 весовых функций приведены сравнения на 15 наборах данных, полученных при импорте из Telegram. Для сравнения весовых функций найдены степенные законы, наиболее приближающие эмпирические распределения полученных весов вершин, оценивается качество полученного приближения. По итогам сделаны выводы о наиболее подходящей весовой функции для построения графа и анализа сети Telegram-каналов в контексте изучения сложных сетей.

Введение

В настоящее время активно исследуется тема построения и анализа графов взаимодействующих объектов, в частности сформированных на данных из социальных сетей и мессенджеров, где вершинами являются пользователи или иные объекты, а ребрами – взаимодействия между ними [1, 2]. В данной работе будет рассмотрен популярный в России мессенджер Telegram, в котором реализована функциональность для организации публичных Telegram-каналов, представляющих собой информационно-новостные ленты сообщений. Многие СМИ, информационные сообщества, блогеры имеют свои Telegram-каналы и регулярно публикуют в них контент, а пользователи Telegram могут подписаться на данные каналы и по-

лучать информацию в виде сообщений от имени канала.

Telegram предоставляет широкий спектр инструментов для ведения каналов. Помимо публикации уникального контента, администраторы канала могут публиковать пост другого канала в своем (делать репост), цитировать или упоминать другие Telegram-каналы, делать ссылки на внешние адреса в сети интернет. Использование данных возможностей создаст связь между Telegram-каналами, что позволяет рассматривать множество каналов и взаимодействий между ними как взвешенный граф взаимодействующих объектов [3].

Изучение таких сетей, сформированных на основе взаимодействий социальных объектов, важно во многих сферах. С помощью анализа можно выявлять источники информации и пути ее распространения, тем самым обеспечивая безопасность в сетях, а также можно настраивать рекомендательные и рекламные системы.

Методы анализа таких сетей предполагают представление исходных данных в виде графа, который в дальнейшем будут анализировать и на основе которого будут делаться выводы о реальных объектах. Поэтому важно выбрать правильную модель построения такого графа.

Построение графов информационного воздействия в сети каналов Telegram

Рассмотрим взвешенный граф $G(V, E)$, у которого V – множество Telegram-каналов, а E – множество ребер, отвечающих за взаимодействия между парами инцидентных им каналов. На данном множестве ребер E определим весовую функцию $w(e_{AB})$, где $e_{AB} \in E$; $A, B \in V$, зависящую от $\delta_{e_{AB}}^U$, $\delta_{e_{AB}}^M$, $\delta_{e_{AB}}^R$. Тут $\delta_{e_{AB}}^U$ – количество общих уникальных внешних ссылок (URL) в постах у пары каналов; $\delta_{e_{AB}}^M$ – количество постов, где канал А упомянул канал В и наоборот; $\delta_{e_{AB}}^R$ – количество репостов канала А сообщений канала В и наоборот. Тогда для функции F , возвращающей неотрицательные значения, получаем модель для построения графа:

$$w(e_{AB}) = F(\delta_{e_{AB}}^U, \delta_{e_{AB}}^M, \delta_{e_{AB}}^R)$$

Описанный в [3] механизм вычисления весов ребер применяется для каждой пары Telegram-каналов из числа вершин графа. Если весовая функция $w(e_{AB})$ для двух каналов равна нулю, то ребра между вершинами нет, если больше нуля, то формируется ребро с посчитанным весом. Таким образом, на основе множества Telegram-каналов V и значений весовой функции $w(e_{AB})$ для каждой пары вершин, формируется множество ребер графа E' . Соответственно получаем взвешенный граф $G(V, E')$ информационного воздействия в сети Telegram-каналов.

Рассматриваемые модели и наборы данных

В зависимости от выбранной функции $F(\delta_{e_{AB}}^U, \delta_{e_{AB}}^M, \delta_{e_{AB}}^R)$ меняется значение весов ребер $w(e_{AB})$ между вершинами-каналами, соответственно

меняется модель и сам граф. Поэтому важно подобрать такую весовую функцию F , чтобы полученная модель учитывала особенности сети каналов Telegram и соответствовала моделям сложных сетей [4, 5].

В рамках данной работы, рассмотрим 12 различных весовых функций $F(\delta_{e_{AB}}^U, \delta_{e_{AB}}^M, \delta_{e_{AB}}^R)$, разделенных на 4 группы и выявим наиболее подходящую для задач анализа сложных сетей:

Линейная группа:

$$F_{1,1} = 1 * \delta_{e_{AB}}^U + 2 * \delta_{e_{AB}}^M + 3 * \delta_{e_{AB}}^R$$

$$F_{1,2} = 1 * \delta_{e_{AB}}^U + 1 * \delta_{e_{AB}}^M + 1 * \delta_{e_{AB}}^R$$

Логарифмическая группа:

$$F_{2,1} = 1 * \ln(\delta_{e_{AB}}^U + 1) + 1 * \ln(\delta_{e_{AB}}^M + 1) + 1 * \ln(\delta_{e_{AB}}^R + 1)$$

$$F_{2,2} = 1 * \ln(\delta_{e_{AB}}^U + 1) + 2 * \ln(\delta_{e_{AB}}^M + 1) + 3 * \ln(\delta_{e_{AB}}^R + 1)$$

Степенная группа:

$$F_{3,1} = 2^{\frac{1 * \delta_{e_{AB}}^U}{\max(\delta_{e_{AB}}^U)}} + 2^{\frac{1 * \delta_{e_{AB}}^M}{\max(\delta_{e_{AB}}^M)}} + 2^{\frac{1 * \delta_{e_{AB}}^R}{\max(\delta_{e_{AB}}^R)}} - 3$$

$$F_{3,2} = 2^{\frac{1 * \delta_{e_{AB}}^U}{\max(\delta_{e_{AB}}^U)}} + 2^{\frac{2 * \delta_{e_{AB}}^M}{\max(\delta_{e_{AB}}^M)}} + 2^{\frac{3 * \delta_{e_{AB}}^R}{\max(\delta_{e_{AB}}^R)}} - 3$$

$$F_{3,3} = 1 * 2^{\frac{1 * \delta_{e_{AB}}^U}{\max(\delta_{e_{AB}}^U)}} + 2 * 2^{\frac{1 * \delta_{e_{AB}}^M}{\max(\delta_{e_{AB}}^M)}} + 3 * 2^{\frac{1 * \delta_{e_{AB}}^R}{\max(\delta_{e_{AB}}^R)}} - 6$$

Смешанная группа:

$$F_{4,1} = \delta_{e_{AB}}^U + \ln(\delta_{e_{AB}}^M + 1) * \ln(\delta_{e_{AB}}^R + 1)$$

$$F_{4,2} = (\delta_{e_{AB}}^U + 1) * \ln(\delta_{e_{AB}}^M + 1) + \ln(\delta_{e_{AB}}^R + 1)$$

$$F_{4,3} = (\delta_{e_{AB}}^U + 1) * \ln(\delta_{e_{AB}}^R + 1) + \ln(\delta_{e_{AB}}^M + 1)$$

$$F_{4,4} = (\delta_{e_{AB}}^U + 1) * (\ln(\delta_{e_{AB}}^M + 1) + \ln(\delta_{e_{AB}}^R + 1))$$

$$F_{4,5} = (\delta_{e_{AB}}^U + 1) * (\ln(\delta_{e_{AB}}^M + 1) * 2 + \ln(\delta_{e_{AB}}^R + 1) * 3)$$

Для сравнения моделей, полученных на основе описанных функций весов, из Telegram при помощи разработанного программного обеспечения, описанного в [3], были скачаны данные по 15 сетям. Полученные графы имеют различный набор каналов-вершин и обладают разными значениями таких ключевых показателей как количество вершин, количество ребер, коэффициенты кластеризации графа и т.п.

Для каждой сети были сформированы по 12 различных взвешенных графов, соответствующих указанным выше моделям. Далее для их сравнения построено распределение весов вершин этих графов (для взвешенного графа весом вершины будем считать сумму весов всех ребер, инцидентных этой вершине) и проведен анализ их характеристик.

Распределение весов вершин графов

В работах по сложным сетям [4-6], полученным на основе взаимодействия утверждается, что распределение весов вершин подчиняется степенному закону: $f(x) = P(X = x) = Cx^{-\alpha}$ с параметром степени α , лежащим в [2; 3]. Поэтому для каждого полученного графа по 15 рассматриваемым сетям будем

приближать полученные распределения весов вершин степенным законом, находить методом максимального правдоподобия параметр степени α и считать соответствующую ошибку – насколько хорошо распределения приближают эмпирические данные.

Для начала убедимся визуально, что полученные распределения весов вершин графов подчиняются степенному закону. Для этого построим функцию плотности вероятности для выбранных графов на логарифмических координатных осях (пример графика на рис.1).

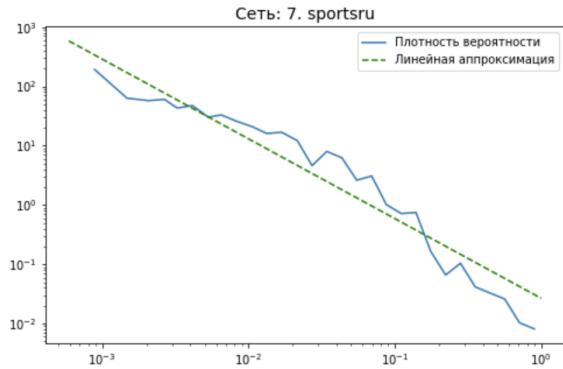


Рис. 1. Плотность вероятности для $F_{1,1}$ и сети, построенной от канала sportsru

Графиком плотности вероятности степенного распределения, построенным на логарифмических осях, является прямая, поэтому аппроксимируем график плотности вероятности для $F_{1,1}$ прямой (зеленая пунктирная линия на рис.1). Как видно на рисунке, значения плотности вероятности хорошо приближаются прямой, поэтому можно сделать вывод, что полученное распределение подчиняется степенному закону.

Сравнение моделей информационного воздействия Telegram

Воспользуемся методом максимального правдоподобия для нахождения параметров степенного закона, наиболее приближающих распределения степеней вершин, такая механика подробно описана в работе [7]. Проведем эту операцию для 15 рассматриваемых сетей и 12 весовых функций, далее сравним полученные результаты. На основе этих данных определим, какая из 12 рассматриваемых весовых функций F лучше подходит для построения графа сложной сети.

В таблице 1 приведены все найденные значения параметра α . Как видно из таблицы, часть значений не попадают в интервал [2, 3] (отмечены оранжевым цветом). А в графах, моделирующих социальные взаимодействия (в частности, социальные сети), распределение весов вершин подчиняется степенному закону с параметром α , попадающим в данный интервал, о чем говорится в работах [4-6].

Таблица 1. Значения параметра α

| | G 1 | G 2 | G 3 | G 4 | G 5 | G 6 | G 7 | G 8 | G 9 | G 10 | G 11 | G 12 | G 13 | G 14 | G 15 |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| F 1.1 | 2,15 | 2,07 | 2,36 | 2,03 | 2,53 | 2,06 | 2,52 | 2,61 | 2,91 | 2,31 | 2,29 | 3,08 | 2,01 | 3,15 | 3,82 |
| F 1.2 | 1,81 | 2,05 | 2,48 | 2,20 | 2,47 | 2,16 | 2,61 | 2,62 | 2,47 | 2,35 | 2,29 | 2,41 | 2,06 | 1,99 | 2,29 |
| F 2.1 | 4,25 | 2,11 | 3,07 | 2,82 | 3,42 | 7,04 | 3,41 | 3,50 | 3,53 | 2,25 | 2,55 | 3,19 | 4,20 | 4,13 | 2,89 |
| F 2.2 | 3,62 | 2,32 | 2,87 | 2,65 | 3,01 | 3,59 | 3,41 | 3,79 | 3,63 | 2,70 | 2,47 | 3,75 | 4,43 | 2,41 | 2,72 |
| F 3.1 | 1,82 | 2,37 | 2,69 | 4,33 | 2,42 | 2,31 | 2,21 | 2,44 | 2,81 | 2,11 | 2,24 | 2,18 | 2,35 | 2,02 | 2,80 |
| F 3.2 | 1,87 | 2,11 | 2,33 | 2,62 | 2,06 | 2,14 | 2,24 | 2,31 | 2,22 | 2,30 | 2,19 | 2,44 | 2,22 | 2,79 | 2,67 |
| F 3.3 | 2,15 | 2,20 | 2,59 | 2,85 | 2,07 | 2,14 | 2,29 | 2,51 | 2,39 | 2,27 | 2,29 | 2,13 | 2,24 | 2,88 | 2,72 |
| F 4.1 | 2,16 | 2,62 | 2,79 | 1,79 | 2,62 | 2,40 | 3,03 | 2,83 | 2,63 | 4,24 | 2,39 | 2,14 | 2,55 | 2,45 | 2,76 |
| F 4.2 | 1,85 | 2,07 | 2,25 | 2,11 | 2,09 | 2,14 | 3,24 | 2,74 | 2,84 | 4,10 | 2,25 | 2,78 | 2,38 | 2,28 | 2,38 |
| F 4.3 | 1,96 | 2,07 | 2,31 | 3,18 | 2,13 | 1,95 | 2,93 | 2,51 | 3,29 | 2,38 | 2,33 | 3,08 | 3,04 | 2,58 | 3,71 |
| F 4.4 | 1,73 | 1,91 | 2,04 | 2,02 | 1,98 | 1,99 | 2,93 | 2,30 | 2,87 | 2,20 | 2,16 | 2,24 | 2,39 | 2,21 | 2,31 |
| F 4.5 | 2,06 | 1,92 | 1,96 | 2,04 | 2,02 | 1,91 | 3,26 | 2,36 | 2,90 | 2,41 | 2,18 | 2,27 | 2,40 | 2,22 | 2,34 |

Проанализировав таблицу подробнее, видно, что не попадание в интервал [2, 3] особенно характерно для логарифмических ($F_{2,1}$, $F_{2,2}$) и смешанных весовых функций ($F_{4,1}$ - $F_{4,5}$). А для рассматриваемых линейных ($F_{1,1}$, $F_{1,2}$) и степенных ($F_{3,1}$, $F_{3,2}$, $F_{3,3}$) функций непопадание в интервал [2, 3] есть, но они значительно реже, в среднем для 1 сети из 10, что является хорошим результатом. Поэтому данные весовые функции F можно рассматривать в роли кандидатов для представления взаимодействий между Telegram-каналами, а логарифмические и смешанные не подходят для этих целей.

Для линейных и степенных моделей посмотрим, насколько хорошо полученные теоретические распределения приближают эмпирические данные. Для этого для каждой сети и весовой функции посчитаем D – расстояние Колмогорова-Смирнова. Посчитанные значения для всех рассматриваемых сетей и весовых функций приведены в таблице 2.

Насколько видно из таблицы 2, первая линейная весовая функция $F_{1,1}$ приближает хуже остальных только данные одной сети. Для остальных сетей результат лучше: для 6 из них расстояние одно из самых минимальных, а для остальных имеет среднее значение.

У второй линейной функции $F_{2,1}$ результат явно хуже: для 6 сетей значение расстояния максимальное из 5 сравниваемых моделей, поэтому данная функция проигрывает первой.

Таблица 2. Значения D , расстояния Колмогорова-Смирнова

| | G 1 | G 2 | G 3 | G 4 | G 5 | G 6 | G 7 | G 8 | G 9 | G 10 | G 11 | G 12 | G 13 | G 14 | G 15 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| F 1.1 | 0,111 | 0,065 | 0,046 | 0,065 | 0,066 | 0,067 | 0,055 | 0,054 | 0,074 | 0,051 | 0,055 | 0,048 | 0,055 | 0,055 | 0,048 |
| F 1.2 | 0,089 | 0,066 | 0,059 | 0,063 | 0,058 | 0,078 | 0,055 | 0,061 | 0,075 | 0,047 | 0,052 | 0,059 | 0,064 | 0,045 | 0,052 |
| F 2.1 | 0,077 | 0,082 | 0,043 | 0,065 | 0,057 | 0,075 | 0,059 | 0,083 | 0,063 | 0,081 | 0,055 | 0,056 | 0,060 | 0,049 | 0,049 |
| F 2.2 | 0,073 | 0,059 | 0,043 | 0,057 | 0,044 | 0,086 | 0,059 | 0,074 | 0,055 | 0,066 | 0,058 | 0,067 | 0,044 | 0,059 | 0,056 |
| F 3.1 | 0,096 | 0,066 | 0,045 | 0,075 | 0,072 | 0,046 | 0,060 | 0,055 | 0,074 | 0,065 | 0,045 | 0,057 | 0,053 | 0,045 | 0,049 |
| F 3.2 | 0,115 | 0,057 | 0,055 | 0,060 | 0,074 | 0,054 | 0,067 | 0,045 | 0,051 | 0,052 | 0,049 | 0,066 | 0,056 | 0,037 | 0,050 |
| F 3.3 | 0,119 | 0,073 | 0,048 | 0,064 | 0,080 | 0,066 | 0,064 | 0,049 | 0,063 | 0,060 | 0,061 | 0,073 | 0,055 | 0,052 | 0,051 |
| F 4.1 | 0,075 | 0,080 | 0,067 | 0,091 | 0,065 | 0,056 | 0,051 | 0,049 | 0,065 | 0,068 | 0,045 | 0,059 | 0,051 | 0,051 | 0,032 |
| F 4.2 | 0,060 | 0,067 | 0,052 | 0,086 | 0,055 | 0,076 | 0,043 | 0,045 | 0,065 | 0,075 | 0,034 | 0,048 | 0,025 | 0,046 | 0,035 |
| F 4.3 | 0,082 | 0,061 | 0,064 | 0,060 | 0,051 | 0,082 | 0,048 | 0,072 | 0,067 | 0,075 | 0,036 | 0,051 | 0,042 | 0,055 | 0,055 |
| F 4.4 | 0,068 | 0,057 | 0,072 | 0,070 | 0,065 | 0,075 | 0,055 | 0,050 | 0,062 | 0,080 | 0,036 | 0,052 | 0,034 | 0,039 | 0,026 |
| F 4.5 | 0,084 | 0,064 | 0,073 | 0,078 | 0,043 | 0,075 | 0,052 | 0,050 | 0,065 | 0,073 | 0,046 | 0,059 | 0,033 | 0,045 | 0,037 |

Далее рассмотрим степенные функции. Функция $F_{3,3}$ явно хуже всего приближает полученные распределения весов вершин: для 9 сетей значение расстояния D наибольшее или одно из наибольших из сравниваемых моделей. Весовая функция $F_{3,2}$ также в большей части случаев (5 из 15) показывает один из худших результатов. А вот функция $F_{3,1}$ показы-

вает неплохой результат. Также, как и первая линейная функция $F_{1,1}$, функция $F_{3,1}$ в 6 случаях показывает один из лучших результатов, но для двух сетей результат наихудший, что склоняет выбор в сторону линейной функции $F_{1,1}$.

Таким образом, учитывая рассчитанный по методу максимального правдоподобия коэффициент степенного распределения α для указанных наборов данных, и посчитанное расстояние Колмогорова-Смирнова, линейная функция $F_{1,1}$ является приоритетной для использования при формировании графов информационного воздействия сети Telegram. Именно эта функция и была использована, например, в работе [8].

Заключение

В рамках данной работы приведено описание сравнения разных моделей для формирования графа информационного воздействия в сети Telegram-каналов. Рассмотрено 12 различных весовых функций, используемых для построения графа и определения веса его ребер. Данные функции были применены для формирования графов на основе импортированных из Telegram данных по 15 сетям. По итогам проведенного сравнения сделан вывод, что для решаемых задач линейная функция $F_{1,1} = 1 * \delta_{eAB}^U + 2 * \delta_{eAB}^M + 3 * \delta_{eAB}^R$, предложенная ранее в работе [8] является приоритетной для использования при формировании графов информационного воздействия в сети Telegram-каналов.

Список литературы

1. Fortunato S., Newman M.E.J. 20 years of network community detection. // Nat. Phys. 2022. 18. P. 848–850.
2. Чеповский А. А. О неявных сообществах на графе взаимодействующих объектов // Успехи кибернетики. 2023. Т.4. № 1. С. 56-64.
3. Попов В.А., Чеповский А.А. Модели импорта данных из мессенджера Telegram // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2022. Т.20. №2. С. 60-71.
4. Newman M. E. J. Networks: An Introduction. Oxford University Press, 2010. 784 p.
5. Fortunato S. Community Detection in Graphs. // Physics Reports. 2010; 486(3):75-174.
6. G. Caldarelli, Scale-Free Networks. Oxford University Press, Oxford (2007). 336pp.
7. Clauset A, Shalizi CR, Newman MEJ Power-law distributions in empirical data. // SIAM Review 51: 661–703 (2009).
8. Попов В. А., Чеповский А. А. Выделение неявных сообществ на графе взаимодействия Telegram-каналов с помощью «метода Галактик». // Труды ИСА РАН. – 2022. – Т.72. – №4. С. 39 – 50.

ИССЛЕДОВАНИЕ ПО АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ ПОЗИТИВНОЙ МОДЕЛИ БЕЗОПАСНОСТИ ДЛЯ MODSECURITY WAF

Вишинёв Е.А.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент прикладной математики
МИЭМ НИУ ВШЭ

Аннотация

В работе изучаются методы автоматического построения позитивной модели безопасности (набор правил безопасности, допускающих только явно разрешённый сетевой трафик и блокирующих остальной сетевой трафик) для широко распространённой технологии «ModSecurity: open-source web application firewall» (ModSecurity: межсетевой экран веб-приложений с открытым исходным кодом) [1]. Автором работы разработан алгоритм, который анализирует сетевой трафик веб-приложения и на основе его количества, дисперсии и ошибок формирует позитивную модель безопасности для ModSecurity. В результате применения алгоритма была повышена защищённость тестируемого веб-приложения от 13 векторов атак, включая атаки, от которых изначально полностью отсутствовала защита в стандартной конфигурации ModSecurity.

Введение

Web application firewall (далее – WAF) – это набор мониторов и фильтров для обнаружения и блокирования сетевых атак на веб-приложения. ModSecurity является проектом с открытым исходным кодом, применяемым в различных системах для защиты веб-приложений [1]. Стандартная конфигурация включает только негативную модель безопасности. Негативная модель безопасности – это совокупность правил, работающих по принципу «разрешено всё, что явно не запрещено». Такой подход неэффективен для защиты от новых и неизвестных атак, так как их невозможно заблокировать известными сигнатурами, и данную проблему можно решить с помощью позитивной модели безопасности. Позитивная модель безопасности – это совокупность правил, работающих по принципу «запрещено всё, что явно не разрешено».

Цель работы – разработать алгоритм генерации позитивной модели безопасности для ModSecurity.

В рамках исследования поставлены следующие задачи:

- 1) определить возможность описания позитивной модели безопасности синтаксисом ModSecurity;
- 2) разработать алгоритм формирования позитивной модели безопасности для ModSecurity;
- 3) применить алгоритм на тестируемом веб-приложении и проверить его защищённость путём имитации различных атак на веб-приложение.

В основной части описаны анализ синтаксиса ModSecurity, авторский алгоритм и его тестирование.

Анализ синтаксиса ModSecurity

Синтаксис ModSecurity позволяет описывать правила проверки HTTP-запросов к веб-приложению, используя следующую конструкцию:

```
SecRule <REQUEST_PATH> "@streq </path> "
"id:<n>,phase:<p>, <actions>",
```

где <REQUEST_PATH> – часть HTTP-запроса, которую анализирует правило, </path> – путь к странице веб-приложения, <n> – идентификатор правила, <p> – номер этапа обработки запроса, <actions> – список действий, которые необходимо выполнить, если запрос соответствует правилу (например, заблокировать или замедлить запрос, объединить правило с другими правилами).

Указанная конструкция позволяет выбирать определённый фрагмент HTTP-запроса и проверять его содержимое с помощью регулярных выражений. Авторский шаблон такого правила выглядит следующим образом:

```
SecRule <REQUEST_PATH> "!@rx <RegEx>"
"phase:1, id: <n>, deny, log, msg: 'CUSTOM AUTO
RULE'",
```

где <REQUEST_PATH> – часть HTTP-запроса, которую анализирует правило, <RegEx> – регулярное выражение, <n> – идентификатор правила, <p> – номер этапа, на котором обрабатывается запрос.

Предложенный автором работы шаблон применяется для формирования правил безопасности веб-приложения авторским алгоритмом.

Алгоритм генерации модели безопасности

Рассмотрим схему алгоритма (рис. 1).



Рис. 1. Алгоритм генерации позитивной модели безопасности

На схеме алгоритма описаны условия проверки в соответствии с формулой (1).

В соответствии со схемой, алгоритм на начальном шаге получает сетевой трафик веб-приложения. Например, сетевой трафик может быть записан в файл, который затем передаётся на вход для алгоритма.

Отметим, что в случае записи трафика в файл необходимо рассмотреть вопрос его безопасного хранения до тех пор, пока алгоритм на его основе не сформирует модель безопасности. В сетевом трафике могут содержаться такие сведения, как персональные данные пользователей или конфиденциальная информация.

Алгоритм получает на вход сетевой трафик, обнаруживает в нём запросы к веб-приложению и распределяет запросы на группы.

Под группой запросов в рамках работы понимается множество запросов, у которых определённый заголовок или тело соответствует одному и тому же сформированному регулярному выражению.

Распределение на группы реализовано для дальнейшего формирования разрешающих правил для тех групп, что определяются алгоритмом как легитимные. Данный подход позволяет однократно собрать сетевой трафик, построить модель безопасности и с её помощью в дальнейшем сразу и без накопления статистики блокировать последующие запросы, если они относятся к той же группе запросов, которая была признана нелегитимной. Это является преимуществом перед подходом, когда критерии проверки применяются не к группе запросов по определённому регулярному выражению, а к конкретному пользователю: данный пользователь будет заблокирован по мере накопления статистики, однако если атака продолжится от другого пользователя, то вредоносные запросы этой же группы блокироваться не будут, пока снова не накопится статистика и новый пользователь не будет заблокирован.

Формирование регулярного выражения, на основе которого происходит разделение запросов на группы, производится последовательным выполнением следующего алгоритма для каждого символа строки:

- Если предыдущий шаг такой же и нет символа + на конце, добавить символ +;
- Если предыдущий шаг такой же и есть символ + на конце, ничего не добавлять;
- Если очередной символ – буква верхнего регистра, то добавить $[A-Z]$;
- Если очередной символ – буква нижнего регистра, то добавить $[a-z]$;
- Если очередной символ – число, то добавить к регулярному выражению $[0-9]$;
- Иначе очередной символ – специальный. Добавить к регулярному выражению $|\mathit{c}$, где c – этот специальный символ.

Пример: строка *Ivanov,1234-5678-9012-3456* переводится в регулярное выражение $[A-Z][a-z]+|[0-9]+|[0-9]+|[0-9]+|[0-9]+$

Группа запросов считается легитимной, если выполнены условия, указанные в формуле (1).

$$\left\{ \begin{array}{l} N_{\text{гр}} > N_0 \\ \frac{N_{\text{гр}}}{N_{\text{общ}}} \times 100\% > n_0\% \\ \frac{s^2_{\text{гр}}}{s^2_{\text{общ}}} > D_0 \\ \frac{M_{\text{гр}}}{N_{\text{гр}}} \times 100\% < M_0\% \end{array} \right. , \quad (1)$$

где N_0 , n_0 , D_0 , M_0 – константы, критические значения, определяемые до выполнения алгоритма. $N_{\text{гр}}$ – число запросов в группе. $N_{\text{общ}}$ – число всех запросов. $s^2_{\text{гр}}$ – дисперсия времени запросов в группе. $s^2_{\text{общ}}$ – дисперсия времени всех запросов. $M_{\text{гр}}$ – число ошибок в группе. Дисперсия времени запросов может быть подсчитана по формуле (2).

$$s^2 = E \left((x_i - E(X))^2 \right), \quad (2)$$

где $E(X) = \frac{\sum_i^n x_i}{n}$ – математическое ожидание выборки значений, X – выборка значений, x_i – элемент выборки значений (в данном случае – время получения запроса), n – количество элементов выборки значений.

Выбор данных критериев проверки обоснован предположением о поведении легитимных пользователей веб-приложения и минимизацией вычислительных ресурсов при проверке данных критериев.

Предположение о поведении легитимных пользователей заключается в том, что легитимного сетевого трафика должно быть значительно больше, чем сетевого трафика от злоумышленников, а, следовательно, вредоносного сетевого трафика должно быть меньше некоторого абсолютного значения и некоторой доли всех запросов. Также распределение вредоносного сетевого трафика по времени может значительно отличаться от распределения легитимного сетевого трафика, и степень отличия распределений может быть подсчитана с помощью их дисперсий в соответствии с формулой (2). Помимо этого, вредоносный сетевой трафик вызывает больше ошибок в веб-приложении, соответственно, если определённая группа запросов вызывает больше ошибок, чем критическое значение, то данную группу запросов можно считать нелегитимной.

Минимизация ресурсов при проверке данных критериев обоснована использованием небольшого количества операций, таких как подсчёт количества элементов массива данных, сравнения, сложения, вычитания, умножения и деления, которое может быть заменено на умножение в противоположной части неравенства – для всех указанных операций вычислительная сложность может не превышать $O(n \log n)$ [2].

При условии выполнения указанных критериев, для группы запросов формируется правило для Mod-Security, разрешающее данную группу запросов. Последним шагом алгоритма является объединение данных разрешающих правил в позитивную модель безопасности. Остальные запросы, не соответствующие

щие прошедшим проверку группам запросов, блокируются.

Тестирование алгоритма

Проверка защищённости веб-приложения реализована с помощью программного обеспечения GoTestWAF [3]. Данное программное обеспечение имитирует атаки на веб-приложение различных векторов и для каждого вектора атак определяет степень защищённости как процентное соотношение заблокированных вредоносных запросов ко всем отправленным вредоносным запросам по данному вектору атак.

Для подтверждения повышения защищённости веб-приложения после применения авторского алгоритма было использовано тестируемое веб-приложение Damn Vulnerable Web Application [4]. Его защищённость была протестирована дважды: до применения модели безопасности, сформированной авторским алгоритмом, то есть со стандартной конфигурацией ModSecurity с правилами безопасности OWASP Core Rule Set [5] (рис.2), и после применения сформированной алгоритмом модели безопасности (рис.3).

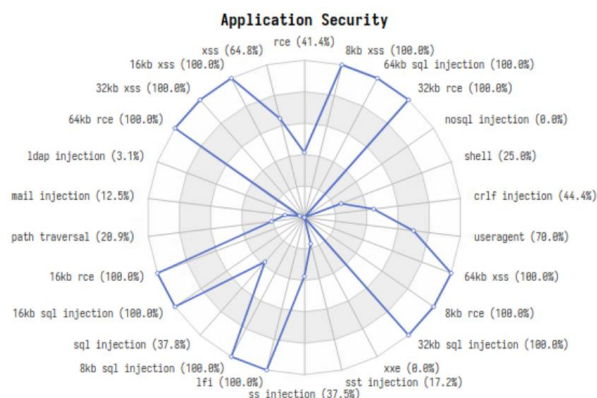


Рис.2. Результаты до применения алгоритма

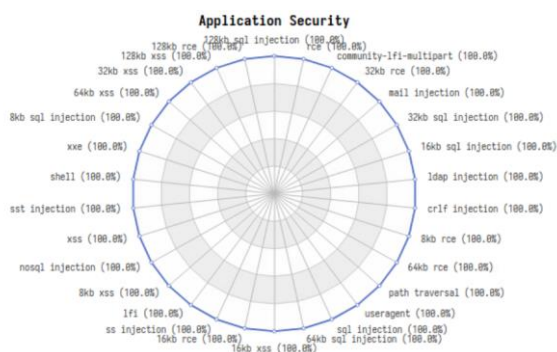


Рис.3. Результаты после применения алгоритма

Применение сформированной модели безопасности повысило степень защищённости веб-приложения от следующих векторов атак:

- Cross-site scripting (XSS) – на 35.2%;
- Remote code execution (RCE) – на 58.6%;
- NoSQL injection – на 100%;
- Shell – на 75%;

- CRLF injection – на 55.6%;
- Useragent – на 30%;
- XML External Entity – на 100%;
- Server Side Template Injection – на 82.8%;
- Server Side Injection – на 62.5%;
- SQL injection – на 62.2%;
- Path traversal – на 79.1%;
- Mail injection – на 87.5%;
- LDAP injection – на 96.9%.

Также заметим, что была сформирована защита от тех векторов атак, для которых изначально отсутствуют правила безопасности в стандартной конфигурации – NoSQL injection и XML External Entity, что подтверждается сравнением результатов защищённости веб-приложения от данных векторов атак. Данный алгоритм способен защитить веб-приложение даже векторов атак, не предусмотренных стандартными правилами защиты, а следовательно – от новых и неизвестных атак.

Заключение

В рамках данного исследования изучены методы формирования позитивной модели безопасности для ModSecurity WAF.

В ходе исследования выяснено следующее:

- 1) синтаксис ModSecurity позволяет формировать позитивную модель безопасности, способную защитить веб-приложения от новых и неизвестных атак;
- 2) предложенный автором алгоритм формирования позитивной модели безопасности построил защиту от 2 векторов атак, для которых отсутствует защита в изначальной конфигурации для ModSecurity;
- 3) авторский алгоритм также повысил защищённость тестируемого веб-приложения от 11 известных векторов атак.

Авторский алгоритм может быть применим для защиты веб-приложений вне зависимости от используемых для их функционирования технологий и содержимого. Ограничения: возможность записи и хранения пользовательского трафика для модели безопасности.

Список литературы

1. Исходный код ModSecurity WAF // Режим доступа: <https://github.com/SpiderLabs/ModSecurity> (дата обращения 13.01.2024).
2. Гашков С. Б. Умножение / Гашков С. Б., Сергеев И. С. - Текст : электронный // Чебышевский сборник. - 2020. - №1 (73). - С. 101-134. - URL: <https://www.chebsbornik.ru/jour/article/view/784> (дата обращения: 20.01.2024).
3. Исходный код GoTestWAF // Режим доступа: <https://github.com/wallarm/gotestwaf> (дата обращения 24.01.2024).
4. Исходный код DVWA // Режим доступа: <https://github.com/digininja/DVWA> (дата обращения 25.01.2024).
5. Официальный сайт OWASP CRS // Режим доступа: <https://coreruleset.org/> (дата обращения 26.01.2024).

СТАНДАРТИЗАЦИЯ ПОСТКВАНТОВЫХ ЦИФРОВЫХ ПОДПИСЕЙ

Сарибекян Г.Э.¹, Валиев Ш.Т.²,
Манжигиевский С.С.², Простенко А.А.¹
Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент прикладной математики,

²департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Постквантовая криптография – современный раздел асимметричной криптографии, посвященный способам построения и анализа алгоритмов, стойких к атакам с помощью как обычных, так и квантовых вычислителей. Алгоритмы постквантовой криптографии и текущие (классические) решения асимметричной криптографии имеют одни и те же интерфейсы. К современным вызовам постквантовой криптографии относят фундаментальные исследования в новых математических областях и выбор наиболее перспективных синтетических решений. Также в данной области требует внимания анализ деятельности международных и национальных организаций по стандартизации и разработка предложений по составу перспективного семейства национальных стандартов в области криптографии с открытым ключом. В работе рассмотрены основные квантово-устойчивые схемы цифровой подписи, выдвинутые за рубежом и в России. Авторами проведено программное тестирование и сравнение данных алгоритмов по ресурсоемкости.

Введение

Угроза квантового компьютера становится все более актуальной. Считается, что алгоритм Шора представляет практическую угрозу для всех классических алгоритмов с открытым ключом. В статье [1] Питер Шор предлагает эффективные рандомизированные алгоритмы разложения чисел на простые множители и дискретного логарифмирования, что говорит о потенциальной угрозе используемым в наше время асимметричным алгоритмам. В качестве асимметричных алгоритмов выступают алгоритмы с открытым ключом, то есть механизмы обмена ключами, схемы формирования и проверки цифровых подписей (ЦП) и алгоритмы асимметричного шифрования. Возможные меры противодействия квантовой угрозе обозначены в [2]:

- Отказ от использования криптографии с открытым ключом при вычислении сеансовых ключей. При таком подходе существуют как минимум 2 проблемы: задача распределения ключей и отсутствие Perfect forward secrecy (PFS). Под PFS понимается свойство некоторых протоколов согласования ключа, которое гарантирует, что сессионные ключи, полученные при помощи набора ключей долговременного пользования, не будут скомпрометированы при компрометации одного из долговременных ключей.

- Применение комбинированных схем с дополнительным симметричным ключом. Суще-

ствуют проблемы с масштабированием, позволяющие противостоять потенциальным угрозам лишь в краткосрочной перспективе.

- Использование квантовой криптографической системы выработки и распределения ключей (ККС ВРК). Несмотря на стойкость и скорость, ККС ВРК подразумевает сложную и дорогую инфраструктуру – использование специального оборудования. На данный момент соединение между двумя точками ограничено 100 км при использовании оптоволоконных линий связи и почти не ограничено по дальности при использовании атмосферных оптических линий связи (АОЛС) [3].

Перспективным решением, решающим перечисленные проблемы, может служить использование новых алгоритмов – постквантовых. Цель работы – провести сравнение зарубежных и отечественных кандидатов на стандартизацию постквантовых цифровых подписей по ресурсоемкости реализаций.

Стандартизация ЦП в США

В 2016 году Национальный институт стандартов и технологий США (NIST USA) объявил конкурс на создание квантово-устойчивых стандартов цифровой подписи и механизма инкапсуляции ключа (далее конкурс NIST). В данной работе сосредоточимся на схемах цифровой подписи. По результатам 3 раунда конкурса NIST было выдвинуто 2 стандарта ЦП:

1. FIPS 204 «Module-Lattice-Based Digital Signature Standard» (ML-DSA) [4].

Стандарт основан на алгоритме *CRYSTALS-Dilithium*, стойкость которого основана на трудоемкости решения задачи LWE (Learning with errors). Для создания ЦП используется модифицированная Байем и Гэлбрейтом техника сжатия цифровой подписи Любашевского [5], которая построена на подходе Фиата-Шамира [6].

2. FIPS 205 «Stateless Hash-Based Digital Signature Standard» (SH-DSA) [7].

Стандарт основан на алгоритме *SPHINCS+*, стойкость которого основана на хэш-функциях. Для создания ЦП используется дерево Меркла с одноразовыми ключевыми парами в листовых узлах FTS (few-time signature) [8].

Сравнительный анализ стандартов цифровой подписи NIST по скорости и памяти

Для проведения тестирования алгоритмов на скорость реализации использовался стенд Ubuntu с ядром 6.5.0-15-generic в Proxmox 8.1.4 с характеристиками, указанными в таблице 1.

Таблица 1. Стенд Ubuntu

| Платформа | Число ядер | Частота, ГГц | Максимальная частота Turbo Boost, ГГц |
|---------------|------------|--------------|---------------------------------------|
| X86-64-v2-AES | 2 | 1,90 | 4,80 |

Реализации постквантовых алгоритмов Dilithium, SPHINCS+ взяты из материалов, предоставленных

авторами на 3 этап конкурса NIST [9]. Обозначим уровень стойкости по NIST как УС.

Таблица 2.
Параметры реализаций Dilithium и SPHINCS+

| Реализация | УС | Размеры параметров, байт | | |
|-----------------------|----|--------------------------|--------------|---------|
| | | Ключ проверки подписи | Ключ подписи | Подпись |
| dilithium | 2 | 1312 | 2032 | 2420 |
| | 3 | 1952 | 2960 | 3293 |
| | 5 | 2592 | 1424 | 4595 |
| sphincs-sha256-128f | 1 | 32 | 64 | 17088 |
| sphincs-sha256-128s | | | | 7856 |
| sphincs-shake256-128f | | | | 17088 |
| sphincs-shake256-128s | | | | 7856 |
| sphincs-sha256-192f | 3 | 48 | 96 | 35664 |
| sphincs-sha256-192s | | | | 16224 |
| sphincs-shake256-192f | | | | 35664 |
| sphincs-shake256-192s | | | | 16224 |
| sphincs-sha256-256f | 5 | 64 | 128 | 49856 |
| sphincs-sha256-256s | | | | 29792 |
| sphincs-sha256-256f | | | | 49856 |
| sphincs-shake256-256s | | | | 29792 |

В Dilithium различные модификации внутри одного уровня стойкости не отличаются друг от друга ни одним из трех параметров, из-за чего в данной таблице объединены.

В SPHINCS+ размеры ключей внутри одного уровня стойкости не отличаются, однако размер подписи sign варьируется. Sign зависит от модификации (f – «быстрая подпись», s – «короткая подпись») и уровня стойкости и не зависит от используемой хэш-функции. Как и следует ожидать, модификации с «короткой» подписью занимают меньший объем памяти по сравнению с «быстрой». При этом их размер даже меньше по сравнению с подписями «быстрых» модификаций уровнем стойкости ниже.

Dilithium и SPHINCS+ отличаются значительно. SPHINCS+ превосходит Dilithium по размерам ключей на 2 порядка. При этом подпись у Dilithium в среднем на 1 порядок короче, даже по сравнению с «короткой» подписью SPHINCS+.

Далее приведены замеры времени генерации ключевой пары (генерация), вычисления подписи (подпись) и проверки подписи (проверка) в мс:

Таблица 3. Замеры времени работы протокольных функций реализаций Dilithium и SPHINCS+

| Реализация | УС | Генерация | Подпись | Проверка |
|-----------------------|----|-----------|----------|----------|
| dilithium2 | 2 | 0,084 | 0,341 | 0,094 |
| dilithium2 AES | | 0,195 | 0,48 | 0,188 |
| dilithium2 AES-R | | 0,19 | 0,533 | 0,183 |
| dilithium2-R | | 0,082 | 0,361 | 0,094 |
| dilithium3 | 3 | 0,159 | 0,573 | 0,155 |
| dilithium3 AES | | 0,346 | 0,837 | 0,316 |
| dilithium3 AES-R | | 0,322 | 0,81 | 0,3 |
| dilithium3-R | | 0,154 | 0,491 | 0,148 |
| dilithium5 | 5 | 0,237 | 0,671 | 0,24 |
| dilithium5 AES | | 0,552 | 1,16 | 0,53 |
| dilithium5-AES-R | | 0,555 | 1,117 | 0,536 |
| dilithium5-R | | 0,227 | 0,703 | 0,238 |
| sphincs-sha256-128f | 1 | 1,302 | 31,354 | 1,785 |
| sphincs-sha256-128s | | 82,364 | 624,476 | 0,62 |
| sphincs-shake256-128f | | 2,166 | 52,565 | 2,916 |
| sphincs-shake256-128s | | 135,783 | 1033,23 | 1,018 |
| sphincs-sha256-192f | 3 | 1,892 | 53,515 | 2,726 |
| sphincs-sha256-192s | | 120,260 | 1150,791 | 0,986 |
| sphincs-shake256-192f | | 3,139 | 85,053 | 4,382 |
| sphincs-shake256-192s | | 199,362 | 1788,881 | 1,488 |
| sphincs-sha256-256f | 5 | 4,876 | 105,172 | 2,735 |
| sphincs-sha256-256s | | 79,516 | 1029,870 | 1,439 |
| sphincs-sha256-256f | | 8,305 | 169,736 | 4,368 |
| sphincs-shake256-256s | | 132,3 | 1578,972 | 2,185 |

Генерация, подпись и проверка для Dilithium отличаются незначительно: за исключением dilithium2 и dilithium5-AES(-R), разница составляет не более одного порядка как между процедурами внутри одной реализации, так и между разными реализациями в целом. При этом в тестировании ни одна из процедур не заняла более 1,2 мс.

Для SPHINCS+ можно выделить три основных тенденции. Во-первых, внутри каждого уровня стойкости реализации, построенные на хэш-функции sha256, производят операции быстрее, чем аналогичные на shake256. Во-вторых, «быстрые» f-реализации выигрывают во времени при генерации ключей и вычислении подписи, но уступают в проверке «ко-

ротким» s-реализациям. В-третьих, «быстрые» f-реализации в среднем требуют время на выполнение процедур, пропорциональное уровню стойкости, что не так для «коротких» s-реализаций.

В сравнении Dilithium и SPHINCS+ между собой, последний уступает по всем временным характеристикам, причем это касается даже сравнения самой устойчивой реализации Dilithium с наименее устойчивой реализацией SPHINCS+.

Стандартизация ЦП в РФ

В 2019 году в рамках 2 подкомитета Технического комитета 26 была создана рабочая группа 2.5 «Постквантовые криптографические механизмы» для решения вопросов стандартизации объектов, относящихся к постквантовым методам криптографической защиты информации. Одна из ключевых задач заключается в разработке стойких постквантовых схем электронной подписи с небольшой длиной ключа и размером подписи [10].

На текущий момент разработаны и представлены научному сообществу алгоритмы, которые планируется подать на стандартизацию:

1) *Шиповник* – схема цифровой подписи, построенная на линейных блочных кодах, исправляющих ошибки. Шиповник содержит набор модификаций цифровой подписи Штерна и преобразования Фиата-Шамира [11].

2) *Гиперикум* - схема цифровой подписи, основанная на мета-схеме SPHINCS+ и отечественной хэш-функции «Стрибог». Доказательство стойкости используется для мета-схемы SPHINCS+ [12]. В отличие от Шиповника схема имеет несколько наборов параметров, что позволяет варьировать скорость и размеры подписей.

Статус валидации ТК 26 предлагаемых постквантовых ЦП [13] приведен в таблице 4.

Таблица 4. Статус валидации регулятором постквантовых ЦП

| ЦП | Статус работ в рамках ТК 26 |
|-----------|---|
| Крыжовник | Реализована первая версия алгоритма. Схема находится на стадии переработки для повышения нефункциональных характеристик. Планируется переработать схему, оформить документы к 4 кварталу 2024 года. |
| Шиповник | В феврале 2023 получен первый пакет замечаний по проекту стандарта. После исправления замечаний обновленный пакет документов по ЦП Шиповник отправлен в апреле 2024 года в ТК 26. |
| Гиперикум | Пакет документов по ЦП Гиперикум отправлен в апреле 2024 года. Готовность спецификации с учетом правок: ориентировочно 2024 год. |

Сравнительный анализ отечественных ЦП по скорости и памяти

Реализации постквантовых криптоалгоритмов Шиповник и Гиперикум взяты из публичного репозитория QApp [14]. Уровень оптимизации хэша «Стрибог-256» (ГОСТ 34.11-2012) выбирался пустым или с использованием инструкций SSE4.1. Для Гиперикума были рассмотрены все доступные наборы параметров: universal, fast_sign, small_sign. Сани-тайзеры не использовались.

Таблица 5. Параметры реализаций Шиповника и Гиперикума

| Реализация | УС | Размеры параметров, байт | | |
|----------------------|----|--------------------------|--------------|---------|
| | | Ключ проверки подписи | Ключ подписи | Подпись |
| Шиповник | 1 | 181 | 362 | 1072662 |
| Гиперикум-universal | 1 | 64 | 128 | 28100 |
| Гиперикум-fast_sign | | | | 58492 |
| Гиперикум-small_sign | | | | 18324 |

Здесь можно отметить, что, хотя Шиповник и Гиперикум базируются на принципиально разных задачах, размеры их ключей отличаются друг от друга не более чем на порядок. При этом подпись Шиповника требует памяти на два порядка больше, чем подпись Гиперикума.

С точки зрения выбора модификации Гиперикума, опция с коротким ключом является более предпочтительной.

Таблица 6. Замеры времени работы протокольных функций реализаций Шиповника и Гиперикума

| Реализация | Генерация | Подпись | Проверка |
|-------------------------------|-----------|--------------|----------|
| Шиповник | 0,841 | 234,731 | 54,330 |
| Шиповник с SSE4.1 | 0,834 | 218,795 | 42,856 |
| Гиперикум | 2480 | 37788 | 34,402 |
| Гиперикум с SSE4.1 | 1599 | 24247 | 22,080 |
| Гиперикум-fast_sign | 79504 | 2767 | 90,960 |
| Гиперикум-fast_sign с SSE4.1 | 49323 | 1723 | 55,921 |
| Гиперикум-small_sign | ~ 20 мин | ~ 1 ч 20 мин | 18,22 |
| Гиперикум-small_sign с SSE4.1 | ~ 10 мин | ~ 1 ч | 11,712 |

В схеме ЦП Шиповник наибольшее время занимает вычисление подписи, на порядок меньше – ее проверка. Еще меньше времени уходит на генерацию ключевой пары: на три порядка относительно подпи-

си. При этом набор инструкций SSE4.1 не оказывает значительного влияния.

На Гиперикум различные модификации оказывают значительное влияние: оптимизированная под SSE4.1 на каждом этапе работает в полтора-два раза быстрее. Ускорение вычисления подписи (*fast_sign*) влечет за собой увеличение времени на генерацию ключей и проверку, причем совокупное время остается сопоставимым. В режиме генерации короткой подписи (*small_sign*) затраченное время не поддается сравнению с остальными режимами и алгоритмами в целом и делает реализацию узкоспециализированной.

Гиперикум показывает несколько лучшие результаты во времени проверки, однако разительно проигрывает по генерации ключевой пары и вычислению подписи, совокупно занимая на 6 порядков больше времени, чем Шиповник.

Заключение

По сравнению с зарубежными отечественные схемы ЦП находятся на более ранней стадии стандартизации. Криптографическому сообществу еще предстоит провести немало исследований как в части стойкости, так и касательно создания ресурсоемких алгоритмов.

После завершения 3 раунда NIST объявил о новом рассмотрении дополнительных кандидатов на стандартизацию в первую очередь для увеличения числа квантово-стойких алгоритмов и появления новых подходов к их синтезу [15]. На 1 раунд нового отбора было принято 40 алгоритмов, из которых 6 основано на кодах, 1 на изогениях, 7 на решетках, 7 на MitM, 10 на многомерных квадратичных системах уравнений, 4 на симметричных схемах и 5 не классифицировано.

Помимо NIST, отечественные коллективы могут изучать несколько менее известные криптографические инициативы, например, такие, как конкурсы Chinese Association for Cryptologic Research (CACR), Korean Post-Quantum Cryptography (Крпс) или рекомендации Bundesamt für Sicherheit in der Informationstechnik (BSI) и др.

Список литературы

1. Shor, Peter W. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer / Peter W. Shor // *SIAM Journal on Computing* : электронный журнал. – URL: <https://epubs.siam.org/doi/10.1137/S0097539795293172> (дата обращения: 15.02.2024). – DOI 10.1137/S0097539795293172
2. Гибридный пост-квантовый обмен ключами в интернет-протоколах // *РусКрипто* : сайт. – URL: https://www.ruscrypto.ru/resource/archive/rc2023/files/02_smyslov.pdf (дата обращения: 17.02.2024)
3. Синергия квантовой и постквантовой криптографии // *РусКрипто* : сайт. – URL: https://www.ruscrypto.ru/resource/archive/rc2023/files/01_fyodorov.pdf?ysclid=ls0qy6ovwh964145962 (дата обращения: 17.02.2024)

4. Module-Lattice-Based Digital Signature Standard // NIST : сайт. – URL: <https://csrc.nist.gov/pubs/fips/204/ipd> (дата обращения: 17.02.2024)

5. An improved compression technique for signatures based on learning with errors // *Cryptology ePrint* : сайт. – URL: <https://eprint.iacr.org/2013/838.pdf> (дата обращения: 17.02.2024)

6. Lattice Signatures without Trapdoors // *SpringerLink* : сайт. – URL: http://link.springer.com/content/pdf/10.1007/978-3-642-29011-4_43.pdf (дата обращения: 17.02.2024)

7. Stateless Hash-Based Digital Signature Standard // NIST : сайт. – URL: <https://csrc.nist.gov/pubs/fips/205/ipd> (дата обращения: 17.02.2024)

8. SPHINCS+ Submission to the 3rd round of the NIST post-quantum project. v3.1 // SPHINCS+ : сайт. – URL: <https://sphincs.org/data/sphincs+-r3.1-specification.pdf> (дата обращения: 13.02.2024)

9. Post-Quantum Cryptography PQC // NIST : сайт. – URL: <https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization/round-3-submissions> (дата обращения: 17.02.2024)

10. Постквантовая криптография. Задачи, перспективы, стандартизация // *Российская академия наук* : сайт. – URL: https://new.ras.ru/upload/iblock/ad0/dnlvvs23wi28ssep_xgm0p4w4ep0u015.pdf (дата обращения: 17.02.2024)

11. О стойкости кодовой электронной подписи на основе протокола идентификации Штерна // *Прикладная дискретная математика* : сайт. – URL: <https://www.mathnet.ru/rus/pdm/y2022/i3/p67> (дата обращения: 17.02.2024)

12. Recovering the tight security proof of SPHINCS+ // *Cryptology ePrint* : сайт. – URL: <https://eprint.iacr.org/2022/346> (дата обращения: 17.02.2024)

13. Квантовая угроза Новые вызовы для инфраструктуры открытых ключей // *PKI Форум России* : сайт. – URL: https://pki-forum.ru/files/files/2023/21_grebnev.pdf (дата обращения: 17.02.2024)

14. GitHub-репозиторий QAPP-tech : сайт. – URL: <https://github.com/QAPP-tech> (дата обращения: 10.02.2024)

15. Post-Quantum Cryptography: Digital Signature Schemes // NIST : сайт. – URL: <https://csrc.nist.gov/Projects/pqc-dig-sig/round-1-additional-signatures> (дата обращения: 17.02.2024)

СХЕМЫ ОБМЕНА КЛЮЧАМИ В ПОСТКВАНТОВОЙ КРИПТОГРАФИИ НА ОСНОВЕ QC-MDPC-КОДОВ

Сарибекян Г.Э.¹, Якушенок К.П.²,
Пашина К.И.², Коробков В.С.²

Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент прикладной математики,

²департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Основанные на теории кодирования криптографические системы, как правило, используют простые двоичные операции и могут быть легко реализованы на программном и аппаратном уровнях. В то же время они требуют больших объемов памяти для хранения ключей, представленных как проверочные матрицы линейных кодов большой размерности. Интересной и актуальной задачей является создание эффективной схемы инкапсуляции ключа на основе задач теории кодирования. В работе рассмотрены алгоритмы помехоустойчивого кодирования постквантовой криптографии. Авторами проведено программное тестирование и сравнение стандарта NIST и схем инкапсуляции ключа на основе QC-MDPC-кодов по ресурсоемкости.

Введение

В 3-м раунде конкурса отбора постквантовых алгоритмов для стандартизации NIST в качестве механизма инкапсуляции ключа был выбран CRYSTALS-KYBER, основанный на теории целочисленных решеток [1], [2]. В настоящее время активно проводятся исследования на стойкость алгоритма [3], [4].

С учетом атак на аппарат изогений суперсингулярных эллиптических кривых [5] криптоалгоритм SIKE был исключен из рассмотрения в 4 этапе конкурса. Актуальными участниками 4-го раунда конкурса NIST стали кандидаты, основанные на кодах, исправляющих ошибки: BIKE, Classic McEliece, HQC [6]. Главным недостатком алгоритма Classic McEliece является большой размер открытого ключа, что является проблемой и для других подобных криптосистем, у которых размер публичного ключа может достигать нескольких сотен мегабайт [7]. Следовательно, возникает потребность в модификации существующих протоколов для обеспечения возможности передачи ключей [8]. В данной работе мы исключаем из рассмотрения Classic McEliece, чтобы избежать указанных проблем. Перейдем к рассмотрению других кандидатов 4 раунда, которые в отличие от криптосистемы Classic McEliece основаны на квази-циклических кодах. Так, в спецификациях алгоритмов BIKE и HQC для редуцирования ключей используют QC-MDPC-коды.

В работе представлен обзор зарубежных алгоритмов обмена ключами, основанных на QC-MDPC-кодах. Также в статье приводится их состоятельность относительно CRYSTALS-KYBER.

Примем за рассмотрение алгоритмов следующую структуру, которую использует NIST [1]: описание, стойкость, эффективность реализаций, изменения после 3 раунда, общая оценка. Для этого введем ряд определений.

Протокол Key Encapsulation Mechanism

Вспомогательные обозначения:

$\lambda \in \{1, 3, 5\}$ - уровень стойкости (УС) соответственно трудоемкости решения задачи нахождения ключа 128, 192, 256-битового блочного шифра.

pk (public key) – открытый ключ.

sk (secret key) – закрытый ключ.

ss (shared secret) – разделяемый секрет (общий ключ).

E_{pk} – функция шифрования ss с помощью pk .

ct (ciphertext) – шифртекст.

D_{sk} – функция расшифрования с помощью sk .

$KeyGen$ – функция генерации пары ключей.

Под $a \leftarrow A$ случайный равновероятный выбор перемешанной a из множества A .

Схема протокола инкапсуляции ключа (Key Encapsulation Mechanism, КЕМ) размера l бит над алфавитом Σ с участием A и B может быть представлена следующим образом:

Генерация ключей:

$$A: KeyGen(\lambda) = (pk, sk)\#$$

$$A \rightarrow B: pk\#$$

Инкапсуляция:

$$B: ss_B \leftarrow \Sigma^l \#$$

$$B: ct = E_{pk}(ss_B)\#$$

$$A \leftarrow B: ct\#$$

Декапсуляция:

$$A: ss_A = D_{sk}(ct)\#$$

Если $ss_B \neq ss_A$, то произошла ошибка при декапсуляции (расшифровании). Иначе протокол закончился успехом.

Согласно [9] протокол Диффи-Хеллмана относится к протоколам выработки общего ключа, так как заранее ни одна из сторон не знает ключ. Напротив, протокол КЕМ можно отнести к протоколам обмена ключами, так как один из абонентов изначально знает разделяемый секрет (общий ключ). Предназначение протокола – передать общий ключ другому абоненту с помощью инкапсуляции открытым ключом и последующей декапсуляции секретным ключом.

Определение QC-MDPC-кода [10]

Определение QC-кода:

Код C называется квази-циклическим или QC-кодом (*quasi-cycle*), если существует некоторое целое n_0 , такое, что всякий циклический сдвиг $c \in C$ на n_0 позиций также является кодовым словом $c' \in C$. Если обозначить $n = n_0 p$, то порождающая и проверочная матрицы кода C являются $p \times p$ циркулянтными блочными матрицами.

Следствие. Для описания порождающей матрицы QC-кода достаточно первой строки, что ведет к экономии памяти при пересылке открытого ключа.

Определение MDPC-кода:

Код C , заданный проверочной матрицей, в которой вес строк равен w , называют MDPC-кодом.

Вспомогательные обозначения:

$t \in N$ – количество вносимых ошибок.

l – размер общего секрета, натуральное число.

r – простое число, для которого 2 является образующим по модулю r .

$w(x)$, $|x|$ – вес Хэмминга вектора x .

F_2^n – конечное поле из алфавита $\{0,1\}$ размерности n .

$\mathcal{R} = F_2[X]/(X^r - 1)$ – фактор кольцо многочленов.

$\mathcal{H}_w = \{(h_0, h_1) \in \mathcal{R}^2 : |h_0| = |h_1| = w/2\}$ – множество закрытых ключей.

\oplus – побитовое исключительное или.

Определение QC-MDPC кода:

Код C длины n с весом строки w , определённый парой проверочных полиномов $(h_0, h_1) \in \mathcal{H}_w$.

Блоки проверочной матрицы $H_0, \dots, H_{n_0-1} \in F_2^{(n-k) \times (n-k)}$ порождены $n - k - 1$ циклическими сдвигами $h_0, \dots, h_{n_0-1} \in F_2^{n-k} : \sum_{i=0}^{n_0-1} |h_i| \leq w$. Проверочная матрица H образуется конкатенацией H_0, \dots, H_{n_0-1} . Порождающая матрица может быть записана в систематическом виде $G = [I_k | Q]$, где

$$Q = \begin{pmatrix} (H_{n_0-1}^{-1} \cdot H_0)^T \\ (H_{n_0-1}^{-1} \cdot H_1)^T \\ \dots \\ (H_{n_0-1}^{-1} \cdot H_{n_0-2})^T \end{pmatrix}.$$

Декодирование QC-MDPC-кода осуществляется с помощью Bit Flipping algorithm [11].

Декодер (*decoder*) принимает на вход вектор s и $(h_0, h_1) \in \mathcal{H}_w$ возвращает $e_0, e_1 \in \mathcal{R}^2$ такие, что $e_0 h_0 + e_1 h_1 = s$ или ошибку декодирования.

VIKE

Для описания алгоритма введем хэш-функции

K, L, H :

$$\begin{aligned} K: F_2^l \times \mathcal{R} \\ \times F_2^l \rightarrow F_2^l, \\ L: \mathcal{R}^2 \rightarrow F_2^l \\ H: F_2^l \end{aligned}$$

Схема VIKE уровня безопасности λ [12]:

Генерация ключей:

Вывод: $(h_0, h_1, \sigma) \in \mathcal{H}_w \times F_2^l, h \in \mathcal{R}$

1. $(h_0, h_1) \xleftarrow{D} \mathcal{H}_w$
2. $h \leftarrow h_1 \cdot h_0^{-1}$
3. $\sigma \xleftarrow{\$} F_2^l$

Инкапсуляция:

Вход: $h \in \mathcal{R}$

Вывод: $ss_B \in K, ct \in \mathcal{R} \times F_2^l$

1. $m \xleftarrow{\$} F_2^l$
2. $e_0, e_1 \leftarrow H(m)$
3. $ct = (e_0 + e_1 h, m \oplus L(e_0, e_1))$
4. $ss_B = K(m, ct)$

Декапсуляция:

Вход: $((h_0, h_1), \sigma) \in \mathcal{H}_w \times F_2^l,$

$c = (ct_0, ct_1) \in \mathcal{R} \times F_2^l$

Вывод: $ss_A \in K$

1. $e' = \text{decoder}(ct_0 h_0, h_0, h_1)$
2. $m' = ct_1 \oplus L(e')$
3. Если $e' = H(m')$, то $ss_A = K(m', ct)$, иначе $ss_A = K(\sigma, ct)$

Алгоритм VIKE (Bit Flipping Key Encapsulation) предложен в 2012 году. В его основе лежат QC-MDPC-коды. VIKE использует двоичный линейный QC-MDPC код и проверочную матрицу для генерации секретных и публичных ключей. Декодер Black-Gray Flip (BGF) используется для расшифрования. По сравнению с 3 раундом была изменена реализация используемой хэш-функции H , чтобы избежать атаки с использованием информации, полученной по побочным каналам утечки.

HQC

Введем дополнительные обозначения:

G, H, K – определяются с помощью SHAKE256 с 512 битами [13].

Encrypt(pk, m):

Вывод: $ct = (u, v)$

1. $e \xleftarrow{\$} \mathcal{R}$ и $w(e) = w_e$
2. $(r_1, r_2) \leftarrow \mathcal{R}^2$ и $w(r_1) = w(r_2) = w_r$
3. $u \leftarrow r_1 + h \cdot r_2$
4. $v \leftarrow m \cdot G + s \cdot r_2 + e$

Decoder(z) – принимает z и декодирует его в соответствии с алгоритмом декодирования.

Decrypt(sk, ct) – возвращает *Decoder*($v - u \cdot y$), где y это часть секретного ключа.

Схема HQC уровня безопасности λ [14]:

Генерация ключа:

Вывод: $seed_{x,y}, seed_h, s$

1. $seed_h, seed_{x,y} \xleftarrow{\$} F_2^\lambda \times F_2^\lambda$
2. $h \xleftarrow{seed_h} F_2^n$,
3. $x, y \xleftarrow{seed_{x,y}} \mathcal{R}^2, |x| = |y| = w$
4. $s \leftarrow x + h \cdot y$

Инкапсуляция:

Вход: $seed_h, s$

Вывод: $ss_B, ct, d, salt$

1. $m \xleftarrow{\$} F_2^k$

2. $salt \stackrel{\$}{\leftarrow} F_2^{128}$
3. $\theta \leftarrow G(m, seed_h, salt)$
4. $ct \leftarrow Encrypt(seed_h, s, m, \theta)$
5. $ss_B \leftarrow K(m, ct)$
6. $d \leftarrow H(m)$

Декапсуляция:

Вход: $seed_{xy}, ct, d, salt$

Вывод: ss_A

1. $m' \leftarrow Decrypt(seed_{xy}, ct)$
2. $\theta' \leftarrow G(m', seed_h, salt)$
3. $ct' \leftarrow Encrypt(seed_h, s, m', \theta')$
4. Если $ct \neq ct'$ или $d \neq H(m')$, то ошибка декодирования, иначе $ss_A = K(m, c)$

Алгоритм HQC (Hamming Quasi-Cyclic) предложен в 2016 году. В его основе также лежат QC-MDPC-коды. HQC использует преимущества квази-циклической структуры, одновременно обеспечивая более прямое сведение безопасности к задаче декодирования случайного линейного кода. Создатели схемы утверждают, что «сложно свести безопасность кодовой схемы к общей задаче декодирования, когда открытый ключ маскирует секретный ключ» [14]. В HQC умение декодировать код не позволяет нарушителю восстановить общий ключ, так как итоговая ошибка за счет конструкции $s=x+y \cdot h$ превосходит корректирующую способность кода. В отличие от BIKE алгоритм HQC не детерминирован, так как на этапе инкапсуляции используются случайные вектора.

Сравнительный анализ алгоритмов

Рассмотрим и сравним алгоритмы BIKE и HQC относительно планируемого стандарта CRYSTALS-Kyber [2]. Под IND-CPA (Chosen-plaintext attack) всюду далее понимаем стойкость к угрозе различения шифртекстов относительно атаки на основе подобранного открытого текста. Под IND-CCA (Chosen-ciphertext attack) понимаем стойкость к угрозе различения шифртекстов относительно атаки на основе подобранного шифртекста.

Каждый из алгоритмов имеет реализацию всех уровней стойкости (таблица 1).

Таблица 1. Стойкость алгоритмов Kyber, BIKE, HQC

| Схема | Уровни стойкости | | | Модель угрозы | |
|-------|------------------|---|---|---------------|---------|
| | 1 | 3 | 5 | IND-CPA | IND-CCA |
| Kyber | + | + | + | + | + |
| BIKE | + | + | + | + | + |
| HQC | + | + | + | + | + |

В данной работе авторами было реализовано тестирование алгоритмов. Реализация CRYSTALS-

KYBER была взята с версии, представленной авторами на 3 раунде [1], реализации BIKE и HQC – с версий, представленных авторами на четвертом этапе конкурса NIST [5].

Для реализации использовался стенд Ubuntu с ядром 6.5.0-15-generic в Proxmox 8.1.4 с характеристиками: платформа – X86-64-v2-AES, число ядер – 2, частота – 1.90 ГГц, максимальная частота Turbo Boost – 4.80 ГГц.

Таблица 2. Параметры реализаций алгоритмов Kyber, BIKE, HQC

| Реализация | УС | Размеры ключей, байт | | |
|------------|----|----------------------|-------|-------|
| | | pk | sk | ct |
| Kyber-512 | 1 | 800 | 1632 | 768 |
| Kyber-768 | 3 | 1184 | 2400 | 1088 |
| Kyber-1024 | 5 | 1568 | 3168 | 1568 |
| BIKE-3114 | 1 | 1541 | 3114 | 1573 |
| BIKE-6198 | 3 | 3083 | 6198 | 3115 |
| BIKE-10276 | 5 | 5122 | 10276 | 5154 |
| HQC-128 | 1 | 2249 | 2289 | 4497 |
| HQC-192 | 3 | 4522 | 4562 | 9042 |
| HQC-256 | 5 | 7245 | 7285 | 14485 |

В таблице 2 приведены размеры ключей для каждой из реализаций. Kyber 5 УС сопоставим с BIKE 1 уровня. HQC в среднем по каждому УС имеет немного большие параметры, чем BIKE.

Для проведения тестирования на эффективность в программный код была добавлена функция, создающая 100 уникальных ключей. Для каждого из этих ключей реализовывались 100 процедур инкапсуляции и декапсуляций.

Сравним время генерации ключевой пары (KeyGen), инкапсуляции (Encaps) и декапсуляции (Decaps) для представленных в таблице 3 реализаций.

Таблица 3. Замеры времени работы протокольных функций реализаций Kyber, BIKE, HQC

| Реализация | УС | Время работы функций, мс | | |
|------------|----|--------------------------|--------|--------|
| | | KeyGen | Encaps | Decaps |
| Kyber-512 | 1 | 0,060 | 0,050 | 0,056 |
| BIKE-3114 | | 2,046 | 0,188 | 16,210 |
| HQC-128 | | 0,239 | 0,505 | 0,815 |
| Kyber-768 | 3 | 0,094 | 0,082 | 0,081 |
| BIKE-6198 | | 8,677 | 0,437 | 34,767 |
| HQC-192 | | 0,687 | 1,426 | 2,205 |
| Kyber-1024 | 5 | 0,086 | 0,078 | 0,085 |
| BIKE-10276 | | 19,080 | 0,923 | 87,616 |
| HQC-256 | | 1,269 | 2,539 | 3,981 |

Примем показатели Kyber за эталон (таблица 3). KeyGen BIKE в 34-221 раза медленнее, HQC в 4-15 раз. Encaps BIKE 5-37 раз дольше, HQC – в 10-35 раз. Decaps BIKE медленнее в 289-1030 раз, HQC – в 14-46 раз. HQC имеет преимущество в декапсуляции за счет того, что в отличие от BIKE знание структуры

кода и самого кода не позволяет злоумышленнику однозначно декодировать передаваемый вектор. Поэтому в HQC есть возможность выбрать код, который быстро декодируется. Напротив, для BIKE это будет слабостью.

Заключение

На данный момент в России не стандартизованы квантово-устойчивые алгоритмы КЕМ и ведутся активные работы над их созданием [16]. В частности, ведется работа над созданием КЕМ на кодах, исправляющих ошибки.

В статье раскрыто одно из направлений построения КЕМ с приемлемым размером публичного ключа – использование кодов квази-циклической структуры. Так, были рассмотрены алгоритмы BIKE и HQC, использующие QC-MDPC-коды. При сравнении данных кодовых КЕМ можно заключить, что, с одной стороны, HQC имеет наилучшие результаты по сравнению с BIKE в скорости генерации ключей и декапсуляции. Также секретный ключ HQC занимает меньше места в памяти. С другой стороны, BIKE имеет достоинства в виде скорости инкапсуляции, размеров открытого ключа и шифртекста.

Список литературы

1. Status Report on the Third Round of the NIST Post-Quantum Cryptography Standardization Process // NIST: сайт. – URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934458 (дата обращения: 29.01.2024)
2. Module-Lattice-Based Key-Encapsulation Mechanism Standard // NIST: сайт. – URL: <https://csrc.nist.gov/pubs/fips/203/ipd> (дата обращения: 29.01.2024)
3. Dubrova, Elena & Ngo, Kalle & Gärtner, Joel & Wang, Ruize. (2023). Breaking a Fifth-Order Masked Implementation of CRYSTALS-Kyber by Copy-Paste. 10-20. 10.1145/3591866.3593072.
4. Уязвимость в реализациях постквантового алгоритма шифрования Kyber // OpenNET: сайт. – URL: <https://www.opennet.ru/opennews/art.shtml?num=60405> (дата обращения: 29.01.2024)
5. An efficient key recovery attack on SIDH // Cryptology: сайт. – URL: <https://eprint.iacr.org/2022/975> (дата обращения: 29.01.2024)
6. Post-Quantum Cryptography // NIST: сайт. – URL: <https://csrc.nist.gov/Projects/post-quantum-cryptography/round-4-submissions> (дата обращения: 29.01.2024)
7. RFC 9242 // Datatracker: сайт. – URL: <https://datatracker.ietf.org/doc/rfc9242/> (дата обращения: 29.01.2024).
8. The McEliece-type Cryptosystem based on D-codes // CTCrypt 2023: сайт. – URL: <https://ctcrypt.ru/files/files/2023/07/11.pdf> (дата обращения: 04.02.2024)
9. Н. В. Никонов, А. В. Черемушкин // Криптографические протоколы, 2022 с. 205–238.
10. Малыгина Е.С. Введение в теорию кодирования // III Математическая студенческая школа НИЯУ МИФИ, г. Москва, 25 ноября 2023 г., Москва, 2023.

С. 39-40.

11. Nilsson, A., Bocharova, I.E., Kudryashov, B.D., & Johansson, T A Weighted Bit Flipping Decoder for QC-MDPC-based Cryptosystems. [Текст] / Nilsson, A., Bocharova, I.E., Kudryashov, B.D., & Johansson, T // IEEE International Symposium on Information Theory (ISIT). — 2021. — № 7. — С. 1266-1271.

12. BIKE - Bit Flipping Key Encapsulation Round 4 Submission // BIKE Website: сайт. – URL: https://bikesuite.org/files/v5.0/BIKE_Spec.2022.10.10.pdf (дата обращения: 16.02.2024)

13. Hamming Quasi-Cyclic (HQC) Fourth round version // HQC: сайт. – URL: https://pqc-hqc.org/doc/hqc-specification_2023-04-30.pdf (дата обращения: 16.02.2024)

14. Квантовая угроза Новые вызовы для инфраструктуры открытых ключей // QApp: сайт. – URL: https://pki-forum.ru/files/files/2023/21_grebnev.pdf (дата обращения: 04.02.2024)

ИССЛЕДОВАНИЕ ПОДХОДОВ К ОРГАНИЗАЦИИ СКРЫТЫХ ВЫЧИСЛЕНИЙ

Павлова Е.Д., Валиев Ш.Т., Коробков В.С., Манжиловский С.С.

*Национальный исследовательский университет «Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Защита конфиденциальности данных стала важной проблемой безопасности при их обработке с использованием облачных платформ. Во многих возможных случаях данные передаются для целей анализа и обработки данных третьим сторонам. Большинство существующих подходов для зашифрованных облачных вычисления построены по уровневым схемам гомоморфного шифрования (Leveled HE, LHE) или схем с некоторой степенью гомоморфности (Somewhat HE, SWHE), обе из которых поддерживают гомоморфные операции, которые позволяют пользователям выполнять вычисления над зашифрованными данными без их расшифрования. Но это возможно только для ограниченного количества операций [1], [2], [3], [4].

Существуют алгоритмы полностью гомоморфного шифрования (FHE), которые потенциально решают эту проблему с помощью процедуры смены ключа (bootstrapping), которая может уменьшить шум зашифрованного текста во время гомоморфных вычислений и, следовательно, позволяет выполнять вычисления для большего количества операций [5].

Введение

Гомоморфное шифрование – позволяет выполнять определенные типы вычислений на зашифрованных текстах и получать результаты вычислений, которые

при расшифровании соответствуют результатам операций, выполняемых на открытом тексте [6].

Алгоритмы гомоморфного шифрования можно разделить на основе вышеуказанных свойств на следующие виды: [7] [17].

Алгоритм шифрования называется частично гомоморфным (Partial Homomorphic Encryption, PHE), если существует эффективный алгоритм, который для любых $x_1, x_2 \in X$ и для любого $k \in K$, получив на вход только $E(k, x_1)$ и $E(k, x_2)$, выдает значение $E(k, x_1 \cdot x_2)$. Данное шифрование позволяет выполнять только один тип операций над зашифрованными данными - либо сложение, либо умножение.

Под ограниченно гомоморфными алгоритмами шифрования (Somewhat Homomorphic Encryption, SWHE) будем подразумевать такие Σ , что для любой функции $f: X^m \rightarrow X$ из некоторого класса функций F и любого $k \in K$ существует эффективный алгоритм, который, получив на вход только $\{E(k, x_i), i = 1, \dots, m\}$, выдает значение $E(k, f(x_1, \dots, x_m))$. Данное шифрование позволяет выполнять более одного типа операций - умножение и сложение, но количество данных вычислений ограничено.

Полностью гомоморфным (Fully Homomorphic Encryption, FHE) алгоритмом шифрования называется, ограниченно гомоморфный алгоритм, в случае, когда F есть множество всех возможных функций $X^m \rightarrow X$ для всех $n \in \mathbb{N}$. Этот алгоритм позволяет выполнять множество операций - умножение и сложение, без ограничения на количество вычислений.

В данной работе мы сосредоточимся на алгоритмах полностью гомоморфного шифрования, так как они позволяют реализовать не одну, а две операции неограниченное количество раз, что делает возможным их применение для более широкого класса задач.

RLWE

Пусть $q \geq 2$ – целое, $R = \mathbb{Z}[x]/(x^d + 1)$ кольцо многочленов над целыми числами, где d – степень 2, $R_q = R/qR$, χ – распределение над R . На вход подаются пары чисел (a_i, b_i) . Задача RLWE заключается в том, чтобы отличить: пара была сформирована случайно как $(a_i, b_i) \in R_q^2$ или сформирована как $a_i \in R_q$, $b_i = a_i \cdot s + e_i$, где $s \in R_q$, $e_i \in \chi$ над R . В LWE используются $(\vec{a}_i, b_i) \in \mathbb{Z}_q^{n+1}$ и $\vec{a}_i \in \mathbb{Z}_q^n$, $b_i = \langle \vec{s}, \vec{a}_i \rangle + e_i \bmod q$ соответственно, где $s \in \mathbb{Z}_q^n$, $e_i \in \chi$ над \mathbb{Z} , $\langle \vec{a}, \vec{b} \rangle = \sum_i \alpha_i \cdot \beta_i$. В модификации GLWE используются $(\vec{a}_i, b_i) \in R_q^{n+1}$ и $\vec{a}_i \in R_q^n$, $b_i = \langle \vec{s}, \vec{a}_i \rangle + e_i \bmod q$: $s \in R_q^n$, $e_i \in \chi$ над R [8].

BGV

BGV – уровневный гомоморфный алгоритм шифрования, стойкость которого основана на трудоёмкости решения задачи RLWE, при этом базовый алгоритм BGV строится на его модификации GLWE [8].

BGV работает с целочисленными сообщениями, представленными в виде векторов в пространстве R_q длиной не более $n + 1$. В процессе вычислений поддерживаются операции сложения и умножения, при этом операция сложения может выполняться быст-

рее, так как реже требует снижать накопленную ошибку e [8].

TFHE

TFHE – «быстрая полностью гомоморфная система шифрования на торе» представляет из себя набор модификаций алгоритма GSW [9], стойкость которого, в свою очередь, основывается на сложности LWE [10].

В качестве входных данных TFHE принимает число из пространства $\mathcal{M} \subseteq T_d[x] = \mathbb{R}[x]/(x^d + 1) \bmod 1$. В качестве гомоморфных операций поддерживаются логические операции: НЕ, И, ИЛИ, И-НЕ, ИЛИ-НЕ, исключающие ИЛИ (XOR), исключающее ИЛИ-НЕ (XNOR), тернарный оператор. Такой набор операций позволяет реализовать приближение любой функции, однако по своей сути является эмуляцией архитектуры процессора, из-за чего неприменим без аппаратной поддержки из соображений скорости работы.

CKKS

Схема CKKS представляет собой гомоморфную криптографическую систему, которая поддерживает приближенную арифметику над комплексными числами \mathbb{C} [11]. Она базируется на трудоёмкости решения задачи RLWE, как и BGV, модификацией которой она является [8]. CKKS поддерживает следующие операции [13]:

1. Сложение,
2. Умножения,
3. Высокоточные CKKS,
4. Масштабирование (автоматическое и ручное),
5. Гибридное переключение клавиш,
6. Подъем.

Его ключевой особенностью является возможность выполнения операций с векторами вещественных чисел \mathbb{R} , что было невозможно делать в крипто-системах Джентри и BGV [14][8].

Применение гомоморфного шифрования в задачах машинного обучения

Вычисления в задачах машинного обучения зачастую являются трудозатратными, в связи с чем производятся в облачных серверах, что приводит к повреждению данных, незаконному вторжению и многим другим рискам безопасности. Гомоморфное шифрование является многообещающим кандидатом для безопасных облачных вычислений, но его применение в реальных задачах машинного обучения является очень сложной задачей.

Далее мы рассмотрим, некоторые алгоритмы машинного обучения, какие операции необходимы для их осуществления и какие схемы гомоморфного шифрования позволяют реализовать данные алгоритмы.

Логистическая регрессия

Данные для обучения в логистической регрессии состоят из пар (x_i, y_i) , где $x_i = (x_{i1}, \dots, x_{id})$ – вектор признаков для обучения, а y_i – метка класса (в дан-

ной работе рассмотрим случай бинарной классификации, где $y_i = 1/-1$. Модель:

$$z = \sigma(y_i \cdot (1, x_i)^T \beta),$$

где $\sigma(x) = 1 / [1 + \exp(-x)]$ – логистическая функция, также называемая сигмоидой, где $\beta = (\beta_0, \beta_1, \dots, \beta_d)$ – параметры модели для обучения. Цель логистической регрессии найти оптимальные значения параметра β , при которых функция ошибки $L(y, z)$ минимизируется:

$$L(y, z) = (1/n) \sum_{1 \leq i \leq n} \log(1 + \exp(-y_i \cdot (1, x_i)^T \beta))$$

При обучении модели градиентным спуском параметры модели обновляются по формуле:

$$\beta \leftarrow \beta + \alpha \frac{1}{n} \sum_{1 \leq i \leq n} \frac{y_i x_i}{1 + \exp(y_i \cdot (1, x_i)^T \beta)}$$

Таким образом, для решения задачи логистической регрессии на полностью зашифрованных данных (шифруются все пары (x_i, y_i)) требуются операции сложения, умножения, деления и функция логарифмирования.

Так как гомоморфное шифрование поддерживает только операции сложения и умножения и лишь некоторые деления, для других функций и операций используют функции аппроксимации.

Линейная регрессия

Данные для обучения в линейной регрессии состоят из пар (x_i, y_i) , где $x_i = (x_{i1}, \dots, x_{id})$ – вектор признаков для обучения, а y_i – целевая переменная. Задача линейной регрессии заключается в нахождении таких весов $\beta = (\beta_0, \beta_1, \dots, \beta_d)$, при которых ошибка $L(y, \beta)$ (рассматривается средняя квадратичная ошибка) будет минимальной:

$$L(y, \beta) = \left(\frac{1}{n}\right) \sum_{1 \leq i \leq n} ((\beta, (1, x_i)) - y_i)^2$$

Найти веса β можно двумя способами. Первый способ – аналитический, в данном случае необходимо решить систему линейных уравнений:

$$\beta = (X^T X)^{-1} X^T y$$

В этом случае (шифруются все пары (x_i, y_i)) потребуются операции умножения и нахождения обратной матрицы [15].

Другой способ нахождения параметров β – с помощью градиентного спуска, где веса модели будут обновляться по формуле:

$$\beta \leftarrow \beta + \alpha \frac{1}{n} \sum_{1 \leq i \leq n} (\beta x_i - y_i) \cdot x_i$$

Для решения задачи линейно регрессии (шифруются все пары (x_i, y_i)) потребуются операции сложения, умножения и вычитания.

K-ближайших соседей (knn)

Мы будем рассматривать knn на Манхэттенском расстоянии, которое является менее трудно реализуемым в вычислениях.

Данные для обучения, состоящие из пар (x_i, y_i) , где x_i представляет собой вектор признаков объекта, а y_i – метка класса объекта x_i , определяющая к какому классу принадлежит объект. Модель:

$$(A, B) = \sum_{i=1}^m |x_i - y_i|$$

Для проведения сортировки зашифрованных данных $cdist'[i]$ в статье [16] предлагается использовать пузырьковую сортировку (bubble sort). Так как при сортировке Манхэттенского расстояния основной операцией является замена двух расстояний на основе их отношения больше-меньше. Следовательно, учитывая два последовательных зашифрованных расстояния $dist[i]'$ с последующим $dist[i+1]'$, затем при сортировке их позиции изменены, если $dist[i]' > dist[i+1]'$.

Для решения задачи k-ближайших соседей потребуются операции сложения, вычитания, модуля и сортировки.

Выводы

В случае логистической регрессии, необходимы операции, которые не поддерживает ни одна из описанных схем (деление, логарифмирование), поэтому используют функции аппроксимации, как, например, в статьях [18].

Для линейной регрессии, в случае аналитического решения требуется вычисление обратной матрицы, что может быть реализовано лишь для ограниченного количества признаков [19]. В случае поиска весов градиентным спуском, все операции возможно реализовать на операциях гомоморфного шифрования [20].

Для реализации k-ближайших соседей возможно использовать алгоритм TFHE, но только на Манхэттенском расстоянии [21] с использованием пузырьковой сортировки и булевой функций AND, OR и XOR. Данный способ не очень оптимален так как слишком трудозатратен. Но в статье [22] описан более оптимальный метод реализации алгоритма k-ближайших соседей основанный на Евклидовом расстоянии и сходстве Пирсона при помощи алгоритма СККС.

Поскольку в машинном обучении принято нормализовать данные перед применением алгоритмов машинного обучения [12], то данные являются нецелыми числами; функции аппроксимации требуют большого числа операций, при которых только растет ошибка [20], СККС является самой подходящей схемой для решения описанных задач, поскольку единственная поддерживает операции сложения и

умножения с нецелыми числами и не требует эмуляции инструкций процессора для их выполнения.

Заключение

Недавние исследования показывают потенциал гомоморфного шифрования для использования в ML алгоритмах, однако для его успешной интеграции необходима оптимизация алгоритмов, адаптация к ограниченным вычислительным ресурсам и обеспечению безопасности при работе с реальными данными и задачами. С учетом активности в исследовательской области и быстрого развития технологий, в будущем, гомоморфное шифрование станет более доступным и эффективным инструментом для обеспечения конфиденциальности данных.

В настоящей статье мы рассмотрели существующие алгоритмы гомоморфного шифрования и его значимость для обеспечения безопасности в машинном обучении. Несмотря на прогресс в этой области, существующие алгоритмы все еще нуждаются в дальнейших исследованиях и улучшениях перед тем, как они смогут быть использованы на практике.

Список литературы

1. CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy // Proceedings of Machine Learning Research: сайт. – URL: <https://proceedings.mlr.press/v48/gilad-bachrach16.html> (дата обращения: 19.02.2024)
2. Optimized Privacy-Preserving CNN Inference With Fully Homomorphic Encryption // Association for Computing Machinery: сайт. – URL: <https://dl.acm.org/doi/abs/10.1109/TIFS.2023.3263631> (дата обращения: 19.02.2024)
3. Efficient CNN building blocks for encrypted data // MBZUAI: сайт. – URL: <https://dclibrary.mbzuai.ac.ae/cgi/viewcontent.cgi?article=1116&context=cvfr> (дата обращения: 19.02.2024)
4. Deep Neural Networks for Encrypted Inference with TFHE // Association for Computing Machinery : сайт. – URL: https://dl.acm.org/doi/abs/10.1007/978-3-031-34671-2_34 (дата обращения: 19.02.2024)
5. Computing on Encrypted Data // SpringerLink: сайт. – URL: https://link.springer.com/chapter/10.1007/978-3-642-10433-6_32 (дата обращения: 19.02.2024)
6. Гомоморфное шифрование: безопасность облачных вычислений и другие приложения (обзор) // Информатика: сайт. – URL: <https://inf.grid.by/jour/article/viewFile/11/13> (дата обращения: 19.02.2024)
7. New Fully Homomorphic Encryption Scheme Based On Multistage Partial Homomorphic Encryption Applied In Cloud Computing // ResearchGate: сайт. – URL: https://www.researchgate.net/publication/339129682_New_Fully_Homomorphic_Encryption_Scheme_Based_On_Multistage_Partial_Homomorphic_Encryption_Applied_In_Cloud_Computing (дата обращения: 19.02.2024)
8. Fully Homomorphic Encryption without Bootstrapping // Cryptology ePrint: сайт. – URL: <https://eprint.iacr.org/2011/277.pdf> (дата обращения: 19.02.2024)
9. TFHE: Fast Fully Homomorphic Encryption over the Torus* // Cryptology ePrint: сайт. – URL: <https://eprint.iacr.org/2018/421.pdf> (дата обращения: 19.02.2024)
10. Homomorphic Encryption from Learning with Errors: Conceptually-Simpler, Asymptotically-Faster, Attribute-Based // Cryptology ePrint : сайт. – URL: <https://eprint.iacr.org/2013/340.pdf> (дата обращения: 19.02.2024)
11. Simpler and Faster BFV Bootstrapping for Arbitrary Plaintext Modulus from CKKS // Cryptology ePrint : сайт. – URL: <https://eprint.iacr.org/2024/109.pdf> (дата обращения: 19.02.2024)
12. Investigating the impact of data normalization on classification performance // ScienceDirect: сайт. – URL: <https://www.sciencedirect.com/science/article/pii/S1568494619302947> (дата обращения: 19.02.2024)
13. OpenFHE: сайт. – URL: [https://openfhe-develop.readthedocs.io/en/latest/sphinx_rsts/intro/quickstart.html?highlight=BGV vectors#simple-mathematical-operations-and-serialization](https://openfhe-develop.readthedocs.io/en/latest/sphinx_rsts/intro/quickstart.html?highlight=BGV%20vectors#simple-mathematical-operations-and-serialization) (дата обращения: 19.02.2024)
14. Fully Homomorphic Encryption Using Ideal Lattices // Carnegie Mellon University: сайт. – URL: <https://www.cs.cmu.edu/~odonnell/hits09/gentry-homomorphic-encryption.pdf> (дата обращения: 19.02.2024)
15. Implementing Linear Regression with Homomorphic Encryption // Science Direct: сайт. – URL: <https://www.sciencedirect.com/science/article/pii/S1877050922005786> (дата обращения: 19.02.2024)
16. Searching and Sorting of Fully Homomorphic Encrypted Data on Cloud // Cryptology ePrint : сайт. – URL: <https://eprint.iacr.org/2015/981> (дата обращения: 19.02.2024)
17. Homomorphic Encryption // ResearchGate: сайт. – URL: https://www.researchgate.net/publication/273846761_Homomorphic_Encryption (дата обращения: 19.02.2024)
18. Scalable and Secure Logistic Regression via Homomorphic Encryption // Cryptology ePrint : сайт. – URL: <https://eprint.iacr.org/2016/111.pdf> (дата обращения: 19.02.2024)
19. Using Homomorphic Encryption for Large Scale Statistical Analysis // The University of Texas at Austin Computer Science: сайт. – URL: https://www.cs.utexas.edu/~dwu4/papers/FHE-SI_Report.pdf (дата обращения: 19.02.2024)
20. Fast and Secure Linear Regression and Biometric Authentication with Security Update // Cryptology ePrint: сайт. – URL: <https://eprint.iacr.org/2015/692> (дата обращения: 19.02.2024)
21. Encrypted Classification Using Secure K-Nearest Neighbour Computation // Springer Link: сайт. – URL: https://link.springer.com/chapter/10.1007/978-3-030-35869-3_13 (дата обращения: 19.02.2024)
22. Secure KNN Classification Scheme Based on Homomorphic Encryption for Cyberspace // Security and Communication Networks: сайт. – URL: <https://www.hindawi.com/journals/scn/2021/8759922/> (дата обращения: 19.02.2024)

ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ПРИ ПРОЕКТИРОВАНИИ И ДОКУМЕНТИРОВАНИИ API С ИСПОЛЬЗОВАНИЕМ OPENAPI И SPECTRAL

Налегач Д.И., Старых В.А.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В данной работе проводится исследование обеспечения информационной безопасности при проектировании и документировании API с использованием OpenAPI и Spectral. Исследование охватывает ключевые аспекты обеспечения безопасности в веб-приложениях и API, рассматривает стандарты проектирования, такие как OpenAPI, и предлагает решение на основе статического анализа с использованием инструмента Spectral.

Введение

В последние годы организации стремятся обеспечить активный переход на архитектуру микросервисов при разработке веб-приложений. Микросервисная архитектура — это архитектурный подход, при котором приложение разбивается на небольшие, автономные и независимые компоненты, называемые микросервисами [1]. Каждый микросервис выполняет определенную функцию или сервис, и может быть разработан, развернут и масштабирован независимо от других микросервисов. Взаимодействие между микросервисами осуществляется через сетевые вызовы, часто с использованием API (Application Programming Interface). Архитектурный стиль REST требует, чтобы связь клиент-сервер не сохраняла состояние. При работе с RESTful API используют понятие CRUD-модели [2], которые документируют с использованием стандарта OpenAPI Specification [3].

Однако, с развитием API возникают новые вызовы в области информационной безопасности. Проектирование безопасных API и их документирование требуют тщательного внимания к деталям, чтобы предотвратить уязвимости, которые могут быть использованы злоумышленниками для атак на веб-приложения и сервисы.

Целью данного исследования является анализ методов обеспечения информационной безопасности при проектировании и документировании API, описанного по стандарту OpenAPI, и инструмента Spectral для статического анализа спецификации API. В данной статье авторы рассматривают проблемы безопасности в веб-приложениях, стандарты безопасности API, предлагают методику по обеспечению безопасности API и фреймворк по валидации спецификации RESTful API с использованием открытой библиотеки Spectral. Такие методы проверки безопасности как фаззинг или использование динамического анализа в данной статье не рассматриваются.

Основная идея заключается в том, чтобы разработать фреймворк статического анализа API, позволяющий обеспечить выполнение требований информационной безопасности еще на этапе описания контракта взаимодействия, не доходя до реализации самого сервиса, что позволит значительно снизить издержки разработки, тестирования и эксплуатации сервиса.

Угроза информационной безопасности API и методы атаки на приложение

Масштаб и взаимозависимость API создают проблемы для анализа безопасности, поскольку выявленные недостатки безопасности в проекте требуют огромных рутинных операций и подвержены ошибкам. Таким образом, за последние годы произошло много утечек данных API. Например, у Venmo [4], T-Mobile [5] и Tesla [6] были недостатки в разработке API, которые привели к раскрытию конфиденциальных данных. Фактически, по оценкам Gartner, к 2025 году API станут наиболее распространенным вектором атак для утечки данных [7]. Кроме того, в результате опроса более 200 специалистов ИТ безопасности отчета Salt Security о безопасности API [8] показывает, что у всех респондентов были проблемы с безопасностью в их API. 46% атак, с которыми они столкнулись, были связаны с недостаточным контролем аутентификации. Обобщенный REST API запрос с потенциальными угрозами безопасности представлен на рис. 1.



Рис. 1. Угрозы безопасности REST API

В данной статье предлагается рассмотреть традиционные и современные методы атак. Анализ этих методов поможет определить уязвимости, с которыми может столкнуться OpenAPI, и предложить эффективные методы защиты.

Традиционные методы атак:

- SQL-инъекции:

Мероприятия по предотвращению: оценка возможности инъекций SQL в параметры запросов API.

- Межсайтовый скриптинг (XSS):

Мероприятия по предотвращению: анализ сценариев, в которых злоумышленник может внедрить вредоносный скрипт в ответ API. Исследование методов смягчения угрозы XSS в контексте веб-приложений, использующих OpenAPI.

- Атаки на аутентификацию и авторизацию:

Мероприятия по предотвращению: проверка наличия уязвимостей в механизмах аутентификации и авторизации. Рассмотрение возможности подбора паролей, форсирования сессий и других методов

нарушения безопасности аутентификации в OpenAPI.

Современные методы атак:

- Атаки на секреты API (API Secrets Attacks):

Мероприятия по предотвращению: изучение современных методов атаки на секреты API, таких как перехват токенов аутентификации. Разработка мер по защите от подобных атак, таких как использование токенов с коротким сроком жизни и использование механизмов обновления токенов.

- Спуфинг API (API Spoofing):

Мероприятия по предотвращению: оценка рисков, связанных с возможностью подделки запросов к API.

- Сетевые атаки и межсетевые атаки (Network Attacks и Man-in-the-Middle Attacks):

Мероприятия по предотвращению: анализ потенциальных угроз, связанных с перехватом и изменением данных, передаваемых между клиентами и серверами API. Разработка средств шифрования и методов предотвращения атак в рамках OpenAPI.

Раздел предоставляет комплексный анализ традиционных и современных методов атак, а также оценку их влияния на безопасность OpenAPI. Полученные результаты послужат основой для разработки эффективных мер по предотвращению угроз и обеспечению безопасности API в контексте проектирования OpenAPI.

Стандарт проектирования OpenAPI

Автоматическое тестирование API с помощью OpenAPI помогает быстро обнаруживать уязвимости и устранять их, что снижает риск взлома системы.

Безопасность данных также обеспечивается поддержкой механизмов безопасности, таких как открытый протокол (схема) авторизации OAuth2 [9]. Анализ ключевых компонентов OpenAPI и их роли в обеспечении безопасности API даёт представление, какие меры предпринимаются для защиты данных и предотвращения угроз (см. рис.2).

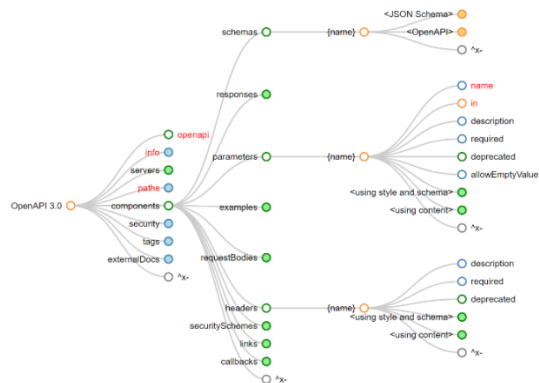


Рис.2. Структура OpenAPI

1. Секция security

Одним из ключевых компонентов является секция security, где определяются механизмы безопасности. Здесь можно указать различные схемы аутентификации и авторизации, такие как API ключи, токены или базовая аутентификация. Четкое определение схем безопасности позволяет контролировать

доступ к ресурсам и предотвращать несанкционированные запросы.

2. Секция components

Секция components предоставляет возможность определить и повторно использовать различные элементы, включая схемы данных, параметры и ответы. Здесь можно определить структуры данных, используемые в API, и установить правила валидации. Это важно для предотвращения атак, связанных с вводом некорректных данных или использованием несанкционированных параметров.

3. Валидация с помощью JSON Schema

Спецификация OpenAPI использует JSON Schema для описания формата данных, передаваемых через API. Использование JSON Schema позволяет валидировать входящие данные, предотвращая атаки, основанные на манипуляции структурой JSON. Это снижает риск внедрения вредоносных данных и обеспечивает целостность передаваемых данных.

4. Секция paths и методы HTTP

Секция paths определяет доступные ресурсы и операции, а методы HTTP определяют действия, которые можно выполнять. Ограничение доступа к некоторым ресурсам или методам может предотвратить нежелательные операции.

5. Секция x-

Кастомные расширения могут быть использованы для настройки механизмов безопасности (например, заголовки или параметры запроса), что позволяет дополнительно защищать API от атак.

6. Секция info

Секция info содержит общую информацию о спецификации API, включая версию, контактные данные и описание. Данный раздел позволит обеспечить своевременное реагирование на инциденты безопасности.

Разработка фреймворка для валидации спецификации OpenAPI

В данном разделе статьи представлено описание процесса разработки инструмента, который обеспечивает валидацию OpenAPI спецификаций на предмет потенциальных угроз безопасности с использованием открытой библиотеки Spectral.

Потребности пользователей:

- Разработчики API нуждаются в инструменте, который поможет обнаружить потенциальные угрозы безопасности в их спецификациях еще на этапе проектирования.

- Разработчики должны иметь возможность настроить правила валидации в соответствии с особенностями и требованиями их проектов.

- Разработчики заинтересованы в интеграции инструмента в их процессы разработки, включая использование CI/CD пайплайнов для автоматической валидации спецификаций.

Название разрабатываемого фреймворка: OpenAPI Security Validator

Функциональность фреймворка:

1. Автоматическая проверка: фреймворк анализирует OpenAPI спецификацию и обнаруживает возможные уязвимости, соответствующие современ-

ным стандартам безопасности - топ 10 Web Application Security Risks (OWASP) [10].

2. Настраиваемые правила: пользователи могут настраивать набор правил проверки в соответствии с требованиями своей системы безопасности.

3. Документация по уязвимостям: фреймворк предоставляет подробные описания обнаруженных уязвимостей, включая их характеристики и рекомендации по исправлению.

4. Интеграция с CI/CD: OpenAPI Security Validator может легко интегрироваться в процесс непрерывной интеграции и доставки (CI/CD), обеспечивая автоматическую проверку безопасности на каждом этапе разработки.

5. Поддержка различных форматов: фреймворк поддерживает различные форматы OpenAPI спецификаций, включая JSON и YAML.

Реализованный фреймворк апробирован на Открытой API проекта Mastercard Open Banking US [11], для поиска потенциальных уязвимостей информационной безопасности. Проект направлен на объединение финансовых данных клиентов с помощью решения Open Banking (Открытый Банкинг), который предоставляется компанией Ficity, входящей в состав Mastercard.

С учетом того, что такие API могут содержать конфиденциальные данные пользователей, платежей и других финансовых данных, то они должны быть стандартизированы, хорошо задокументированы и отвечать всем требованиям информационной безопасности.

Предварительная обработка данных, общие характеристики:

- спецификация API стандартизирована и описана по формату OpenAPI 3.0.3;
- наименование API - Open Banking;
- формат документа YAML;
- содержит 23 405 строк;
- подключение производится к хосту <https://api.fincity.com>;
- содержит 26 разделов (тегов);
- версия спецификации 1.16.1.

Статистика о вызовах API внутри спецификации API:

- 108 вызова API;
- 93 уникальных ресурсов;
- 506 схем запросов и ответов.

Ручной анализ показывает, что имена полей, такие как id, name, description, number и value, часто повторяются в схемах. Однако их значения зависят от контекста, в отличие от более конкретных имен полей, таких как PayrollAccountId.

Было обнаружено, что для всех API определен объект безопасности, а это означает, что для доступа к API можно использовать соответствующие методы аутентификации: FincityAppKey или FincityAppToken. В то же время в описании многих вызовов API не упоминается схема аутентификации и ее обязательность, что указывает на то, что вызов API может быть выполнен без прохождения процесса аутентификации. Следовательно, анализируется объект описания, чтобы определить, является ли аутен-

тификация обязательной для вызова API. В частности, считается аутентификацию вызова API обязательной, только если находится явное «обязательное» требование, указанное в его описании OpenAPI. Используя эти допущения идентифицировано 51 вызов API с явными метками обязательной аутентификации.

По результатам статического анализа спецификации Открытой API проекта Mastercard Open Banking US [11] было обнаружено 1866 проблем. Из них 946 error, 919 warnings и 1 info.

Подробнее об обнаруженных ошибках описано в Таблице 1.

Таблица 1.
Уязвимости Открытой API

| № | Мнемокод правила валидации | Количество обнаруженных | Класс критичности |
|----|--|-------------------------|-------------------|
| 1 | owasp:api1:2023-no-numeric-ids | 4 | Error |
| 2 | owasp:api4:2023-rate-limit-responses-429 | 201 | Warning |
| 3 | owasp:api4:2023-rate-limit | 109 | Error |
| 4 | owasp:api4:2023-string-limit | 445 | Error |
| 5 | owasp:api4:2023-string-restricted | 495 | Error |
| 6 | owasp:api4:2023-array-limit | 148 | Error |
| 7 | owasp:api4:2023-integer-limit-legacy | 194 | Error |
| 8 | owasp:api4:2023-integer-format | 43 | Error |
| 9 | owasp:api7:2023-concerning-url-parameter | 1 | Info |
| 10 | owasp:api8:2023-define-error-responses-500 | 208 | Warning |
| 11 | owasp:api8:2023-define-error-validation | 7 | Warning |
| 12 | owasp:api8:2023-define-cors-origin | 2 | Error |
| 13 | owasp:api8:2023-define-error-responses-401 | 8 | Warning |
| 14 | owasp:api9:2023-inventory-access | 1 | Error |

Вывод: Открытая API проекта Open Banking содержит ряд потенциальных угроз информационной безопасности, что несет за собой риски и может привести к инциденту ИБ. Наиболее распространенные угрозы безопасности API4:2023 - Unrestricted Resource Consumption и API8:2023 — Security Misconfiguration

Пример выходного результата валидации с перечнем ошибок валидации представлен на рис.3.

```

23016:27 error owasp:api4:2023-array-limit components.schemas.VOIEtWierifyReportIncomeStream.properties.transactions
Schema of type array must specify maxItems.
23032:18 error owasp:api4:2023-array-limit components.schemas.VOIEtWierifyInterviewData.properties.tWierifyInterview
Schema of type array must specify maxItems.
23084:18 error owasp:api4:2023-integer-limit-legacy components.schemas.VOIEtWierifyStatementData.properties.assetIds
Schema of type integer must specify minimum and maximum.
23093:26 error owasp:api4:2023-array-limit components.schemas.VOIEtWierifyReport.allOf[1].properties.institutions
Schema of type array must specify maxItems.
23098:20 error owasp:api4:2023-array-limit components.schemas.VOIEtWierifyReport.allOf[1].properties.income
Schema of type array must specify maxItems.

```

Рис.3. Пример результата валидации

Методика обеспечения безопасности API

В этом разделе предлагается рассмотреть методику обеспечения безопасности API с использованием OpenAPI и Spectral.

1. Стандартизация и использование OpenAPI 3.1 при проектировании спецификации API;
2. Внедрять многократную аутентификацию для повышения уровня безопасности;
3. Использовать механизмы для обнаружения и блокировки атак на аутентификацию, такие как блокировка аккаунта после нескольких неудачных попыток. Внедрять мониторинг сессий и токенов для обнаружения подозрительной активности;
4. Использовать параметризованные запросы для предотвращения атак SQL-инъекций и применять фильтрацию и валидацию входных данных;
5. Использовать протокол HTTPS, внедрять механизмы защиты от атак Man-in-the-Middle с помощью Certificate Pinning;
6. Обеспечить механизмы мониторинга и аудита безопасности;
7. Внедрить ограничение скорости запросов, использовать Rate-Limits;
8. Регулярное обновление процедур безопасности и политики, обучение персонала;
9. Интегрировать проверки безопасности в процессы CI/CD для обеспечения непрерывной безопасности;
10. Сотрудничество с сообществом и информационное обмена.
11. И в заключении, использовать предлагаемый фреймворк для валидации разрабатываемого API на этапе проектирования спецификации.

Заключение

Разработанный фреймворк для валидации OpenAPI спецификаций на наличие потенциальных угроз безопасности на основе инструмента Spectral позволяет обеспечить безопасность при проектировании и использовании API. Он обеспечивает поддержку как командной строки, так и API для интеграции с CI/CD системами, что делает его удобным инструментом для внедрения в процессы разработки и обеспечения безопасности.

Данная работа вносит вклад в область информационной безопасности разработки программного обеспечения, предоставляя практические рекомендации и инструменты для обеспечения безопасности при работе с API. Дальнейшие исследования могут быть направлены на расширение функциональности фреймворка, а также на анализ новых угроз и разработку соответствующих методов защиты.

Список литературы

1. R. Fielding, “Architectural styles and the design of network-based software architectures,” Ph.D. dissertation, University of California, Irvine, 2000.
2. Рыбанов А.А. CRUD-модель безопасности как базовый элемент контроля и управления доступом к данным в информационных системах // Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. 2022. №4 (311). URL: <https://cyberleninka.ru/article/n/crud-model-bezopasnosti-kak-bazovyy-element-kontrolya-i-upravleniya-dostupom-k-dannym-v-informatsionnyh-sistemah> (дата обращения: 17.02.2024).
3. “OpenAPI Specification - Version 3.0.3,” Feb 2020. [Online]. Available: <https://swagger.io/specification/>
4. D. Salmon. (2019) I scraped millions of venmo payments, your data is at risk. [Online]. Available: <https://www.wired.com/story/i-scraped-millions-of-venmo-payments-your-data-is-at-risk/>
5. T. Spring. (2018) T-mobile alerts 2.3 million customers of data breach tied to leaky api. [Online]. Available: <https://threatpost.com/t-mobile-alerts-2-3-million-customers-of-data-breach-tied-to-leaky-api/136896/>
6. D. Abdine. (2020) Tesla vulnerability disclosure: Tesla backup gateways. [Online]. Available: <https://www.rapid7.com/blog/post/2020/11/17/dont-put-it-on-the-internet-tesla-backup-gateway-edition/>
7. M. O’Neill, D. Zumerle, and J. D’Hoinne, “How to Build an Effective API Security Strategy,” Gartner Research, Tech. Rep., Dec 2017.
8. “State of API Security,” Salt Security, Tech. Rep., Jan 2021.
9. Юрина Е.С. Authorization in web application under OAuth2.0 protocol // Вестник науки. 2018. №8 (8). URL: <https://cyberleninka.ru/article/n/authorization-in-web-application-under-oauth-2-0-protocol> (дата обращения: 18.02.2024).
10. OWASP Foundation, the Open Source Foundation for Application Security [В Интернете] // OWASP. - 2023 г.. - <https://owasp.org/>.
11. Mastercard OpenBanking US Project [В Интернете] // Mastercard - <https://developer.mastercard.com/open-banking-us/documentation/onboarding/>

ПРОГРАММИРУЕМЫЕ ОТРАЖАЮЩИЕ ПОВЕРХНОСТИ ДЛЯ БЕСПРОВОДНОЙ СВЯЗИ НОВОГО ПОКОЛЕНИЯ

Львов А.В., Иващенко И.В.,
Бондарева П.И., Панарина П.В.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В настоящее время проводится множество исследований в области новых решений для беспроводных сетей связи 5 и 6 поколения. Одним из таких решений являются программируемые отражающие поверхности (ПОП). В данной работе рассматриваются разработка и исследование фазово сконфигурированного линейно поляризованного образца ПОП с оригинальным дизайном фазовращателя.

Введение

Изучение инновационных беспроводных технологий является актуальной областью исследований и разработок. Применение сетей связи нового поколения из-за атмосферных потерь при распространении излучения [1] предъявляет особые требования к формированию узконаправленных пучков излучения [2]. Этот факт ставит перед исследователем задачу создания антенн с узкой диаграммой направленности, с возможностью управления и перестройки диаграммы направленности в реальном времени. Новые аппаратные и программные решения призваны уменьшить потери на распространение сигнала из-за макро- и микромобильности, а также динамических блокировок сигнала [3]. В ряде случаев наиболее оптимальным решением являются ПОП.

ПОП представляет собой пассивную рефлекторную антенну, которая состоит из массива патч-антенн и фазовращателя. Фазовращатель в свою очередь оснащен системой управления набегом фазы, что позволяет задавать направление отражения падающего излучения. Это решение потенциально позволяет обеспечить стабильность сигнала путём размещения сети ПОП в виде аппаратно-программного комплекса, отслеживающего состояние канала связи между базовой станцией и пользовательским оборудованием [4].

Научным коллективом базовой кафедры квантовой оптики и телекоммуникаций ЗАО «Сконтел» МИЭМ ВШЭ ранее проводились научные исследования схожей тематики в диапазоне частот 5/6G. Актуальность данной работы связана с отсутствием коммерчески доступных устройств в диапазоне частот 5G. Данная работа сконцентрирована на исследовании фазово сконфигурированного экспериментального образца на указанный диапазон частот.

Электромагнитная модель

В ходе аналитического обзора научно-технической информации были проанализированы различные решения в области ПОП для частот 2,7 и 15 ГГц [4, 5]. На основе приведенных данных была разработана электромагнитная (ЭМ) модель структуры ПОП в среде численного моделирования, которая использует метод конечных элементов для численного расчета. Структура модели приведена на рис.1.

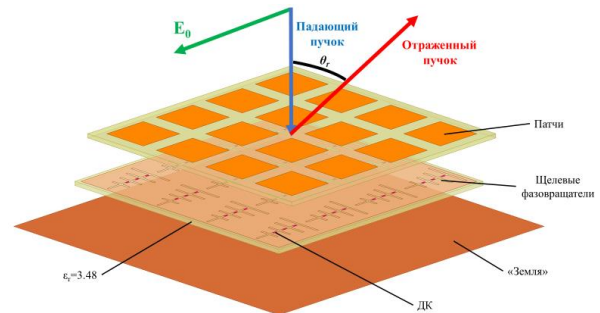


Рис.1. Схема устройства ПОП

Управление конфигурацией ПОП осуществляется при помощи набора диодных ключей (ДК), которые задают набег фаз в ячейках ПОП. Для упрощения конструкции при расчетах и прототипировании каждый диод представлен либо проходным участком щелевой линии (диод в состоянии «закрыто»), либо закороченным участком щелевой линии (диод в состоянии «открыто»).

По результатам моделирования получен профиль Гауссова пучка в отражении, демонстрирующий отклонение от законов геометрической оптики при отражении от ПОП. В качестве основной характеристики модели ПОП была построена диаграмма направленности отражателя в плоскости отражения (рис.2). Ожидаемый угол отклонения пучка, отраженного от экспериментального образца ПОП, определенный по максимуму отраженного излучения, согласно приведенной на графике диаграмме составляет $\theta_r = 62,5^\circ$.

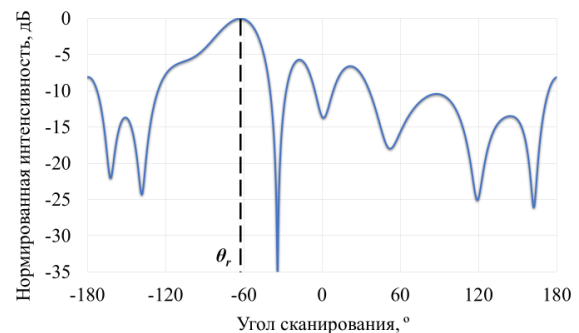


Рис.2. Профиль отраженного от ПОП пучка в E-плоскости, полученный по результатам моделирования

Материалы и методы

В качестве материальной основы для экспериментального образца ПОП был выбран фольгированный диэлектрик FSD888T со следующими характеристиками: толщина 1,524 мм, диэлектрическая проницаемость $\epsilon_r = 3,48$, тангенс угла диэлектрических потерь $\tan \delta = 0,0012$. Данный материал выбирался из соображений достаточной механической прочности и низких электрических потерь.

Методом изготовления экспериментального образца ПОП была выбрана технология фрезерования печатных плат с использованием станка и числовым программным управлением, как оптимально совместимая с толщиной фольгированного диэлектрика и геометрическими параметрами структуры. Технологическая схема представлена на рис.3, где серым обозначены области поверхностного фрезерования для снятия металлизации (серым) и сквозного фрезерования (бирюзовым).

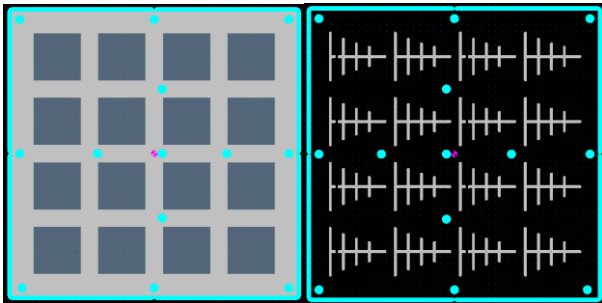


Рис.3. Технологические схемы обработки материала для ЧПУ станка

Экспериментальная характеристика изготовленного образца произведена методом измерения диаграммы направленности, который позволяет получить исчерпывающую информацию о способности образца ПОП производить отражение падающего СВЧ пучка под заданным углом.

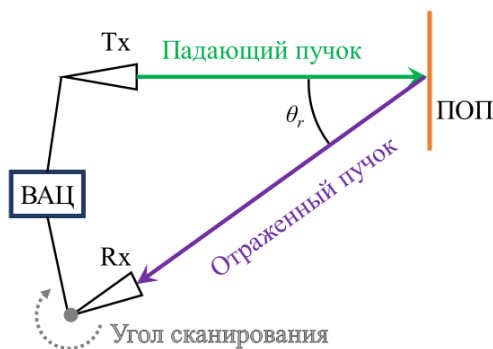


Рис.4. Схема экспериментальной установки

В соответствии со схемой экспериментальной установки, представленной ранее на рис.4, образец ПОП был помещен на линию не прямой видимости между Tx и Rx для проверки его отражательных свойств. Принимая во внимание результаты расчетов, мы опираемся на значение $\theta_r = 62,5^\circ$ при ориен-

тации оптической оси Rx, при этом Tx светит на ПОП перпендикулярно поверхности.

В качестве Tx и Rx использовались антенны на основе планарных массивов патч-антенн 1x4, оснащенных фидерной линией, для которых известна диаграмма направленности. Антенны изготавливались таким же методом на том же материале. В процессе измерений Tx и прототип ПОП находились в статическом положении, а оптическая ось Rx разворачивалась от -45° до $+45^\circ$ с шагом $0,5-1^\circ$. Tx и Rx в свою очередь подключены к векторному анализатору цепей (ВАЦ), при помощи которого мы получаем значения комплексного коэффициента передачи.

Геометрическое расположение элементов антенной сборки в процессе измерений было следующим:

- расстояние Tx-ПОП $L_{12} = 1,4$ м;
- расстояние ПОП-Rx $L_{23} = 1,27$ м;
- угол отклонения пучка $\theta_r = 46,5^\circ$.

Обсуждение результатов

Результаты измерения модуля S_{21} в децибелах для изготовленного образца ПОП представлены на рис.5.

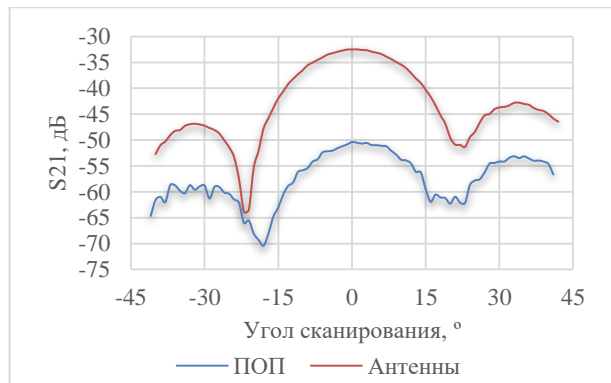


Рис.5. Измеренный профиль отраженного пучка в E-плоскости на частоте 4,8 ГГц

При построении профилей отраженного от ПОП пучка и пучков антенн Tx и Rx были учтены потери на увеличение оптического пути в тракте с использованием уравнения передачи Фрииса:

$$\frac{P_{Rx}}{P_{Tx}} = G_{Tx} G_{Rx} \left(\frac{\lambda}{4\pi L} \right)^2, \quad (1)$$

где P_{Rx} – принятая мощность, P_{Tx} – переданная мощность, G_{Rx} и G_{Tx} – коэффициенты усиления антенн, λ – длина волны, L – расстояние между антеннами. При измерении диаграммы направленности антенны расстояние Tx-Rx составляло 1,4 м, а полная длина оптического пути при измерении диаграммы направленности ПОП составляла 2,67 м. Таким образом, мы получили калибровочный коэффициент 5,6 дБ, который был использован для корректировки измеренных значений модуля S_{21} для профилей пучков антенн.

Анализ экспериментальных данных показал следующие ключевые результаты:

- Рабочая частота 4,8 ГГц,
- Угол отклонения пучка 46,5°.

Таким образом, результаты эксперимента показывают работоспособность системы, которая демонстрирует отклонение падающего пучка от законов геометрической оптики. Отклонения полученных результатов объясняются наличием i) технологических погрешностей при изготовлении антенн и щелевых фазовращателей фрезерной обработкой возможны отклонения от номиналов линейных размеров до $\Delta x = 50$ мкм, а также ii) особенностей режимов работы ЧПУ станка.

Заключение

В данной работе представлены результаты моделирования, подход к изготовлению, методика измерений и анализ результатов экспериментальных измерений фазово сконфигурированного линейно поляризованного образца ПОП с оригинальным дизайном фазовращателя. Согласно результатам моделирования образца ПОП, выполненного в среде ЭМ моделирования методом конечных элементов, ожидается, что угол отклонения нормально падающего СВЧ пучка от ПОП составит $\theta_r = 62,5^\circ$.

Измерение диаграммы направленности образца, изготовленного по технологии фрезерования печатных плат, показало качественное соответствие расчетного и экспериментально полученного угла отклонения падающего СВЧ пучка. При этом погрешности изготовления привели к отклонению параметров от модельных на 25% по углу отклонения и на 4% от рабочей частоты, что позволяет утверждать принципиальную работоспособность системы и дает вектор дальнейших исследований.

В дальнейшем возможно исследование многопозиционной ПОП использующей набор диодных ключей в качестве управляющей системы. Сетка из нескольких таких устройств может быть использована для сверхбыстрой маршрутизации СВЧ пучка при возникновении динамических блокировок радиосигнала в беспроводном канале связи 5G

Благодарности

Авторы статьи выражают благодарность за консультации и активное участие в анализе результатов на всех этапах исследования сотрудникам базовой кафедры квантовой оптики и телекоммуникаций ЗАО «Сконтел» МИЭМ ВШЭ: старшему научному сотруднику Шуракову Александру Сергеевичу и доценту Ожегову Роману Викторовичу.

Список литературы

1. Jornet J. M., Akyildiz I. F. Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band //IEEE Transactions on Wireless Communications. – 2011. – Т. 10. – №. 10. – С. 3211-3221.
2. Akyildiz I. F., Jornet J. M. Realizing ultra-massive MIMO (1024× 1024) communication in the (0.06–10) terahertz band //Nano Communication Networks. – 2016. – Т. 8. – С. 46-54.

3. Shurakov A. et al. Empirical blockage characterization and detection in indoor sub-THz communications //Computer Communications. – 2023. – Т. 201. – С. 48-58.

4. Dai L. et al. Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results //IEEE access. – 2020. – Т. 8. – С. 45913-45923.

5. A. V. Lvov, A. N. Prikhodko, A. S. Shurakov, G. N. Goltsman. Reconfigurable Reflectarrays for 5/6G Wireless Systems with Linear Polarization. SaintPetersburg OPEN.

ИССЛЕДОВАНИЕ ВОЛНОВОДНОГО ПОЛЯРИЗАТОРА С КЛИНОВИДНОЙ МАГНИТНОЙ СТЕНКОЙ

Машкова М.А.

*Московский технический университет
связи и информатики,
кафедра «Электроника»*

Аннотация

В работе проводится компьютерное моделирование волноводного поляризатора на основе клиновидной магнитной стенки из грибовидного метаматериала. С помощью программных средств CST Studio Suite получены диаграммы направленности излучения из открытого конца поляризатора на различных частотах.

Введение

Поляризация распространяющейся электромагнитной волны играет одну из важнейших ролей в радиосвязи, спутниковой связи, радиолокации и других областях науки и техники, позволяя эффективно принимать радиосигналы при известном законе изменения вектора электрического поля в пространстве. И современные темпы развития и модернизации микроволновых устройств, в том числе и поляризаторов электромагнитной волны, приводят к росту требований не только к их электродинамическим характеристикам, надежности, но и массогабаритным параметрам, простоте монтажа, прочности и износостойкости. Актуальность этой проблемы подтверждает ряд публикаций и патентов [1-4], посвященных разработке, компьютерному моделированию и экспериментальному исследованию различных моделей поляризаторов.

Одной из тенденций развития современных микроволновых устройств является внедрение в них структур на основе метаматериалов, что позволит добиться улучшенных электродинамических и массогабаритных характеристик.

В работе предлагается модель волноводного поляризатора, выполненная на основе отрезка прямоугольного волновода типа WR-284 с расположенной в нем клиновидной магнитной стенкой из грибовидного метаматериала. В ходе компьютерного модели-

рования получены диаграммы направленности из открытого конца волноводной конструкции в диапазоне частот от 600 МГц до 4 ГГц, демонстрирующие изменение поляризации распространяющейся электромагнитной волны в зависимости от выбранной частоты.

Поляризация электромагнитной волны

У бегущей плоской электромагнитной волны векторы электрической и магнитной напряженностей лежат в плоскостях, перпендикулярных друг другу, и образуют с волновым вектором \vec{k} правую тройку векторов. Это следует из решений уравнений Максвелла. Из условия однозначности связности векторов \vec{E} и \vec{H} для описания поляризации достаточно описать поведение лишь одного из них. Для этой цели как правило выбирается вектор напряженности электрического поля. Если направление и относительная амплитуда вектора \vec{E} изменяются хаотично с течением времени, то электромагнитную волну принято называть естественной или неполяризованной. У поляризованной электромагнитной волны с течением времени изменение пространственной ориентации вектора \vec{E} происходит по определенному закону.

Поляризация есть свойство электромагнитной волны, характеризующее изменение во времени направления и относительной амплитуды вектора электрического поля [5]. Поляризацию можно представить в виде кривой, которую описывает конец этого вектора в фиксированной точке пространства как функцию времени при наблюдении в направлении распространения волны (см. рис. 1).



Рис. 1. Поведение вектора напряженности электрического поля при различных видах поляризации

Явление эллиптической (круговой) поляризации электромагнитной волны можно описать как движение вектора \vec{E} по замкнутой эллиптической (в частном случае круговой) траектории в плоскости, перпендикулярной направлению распространения волны. В случае, когда конец вектора \vec{E} лежит на прямой линии, имеет место линейная или плоская поляризация. Различают линейные горизонтальную и вертикальную поляризации.

Волноводные поляризаторы

Поляризатор – устройство, предназначенное для преобразования одного вида поляризации электромагнитной волны в другой. Существует возможность создания таких устройств на основе отрезков волноводов различных сечений. Так, размещение диэлектрической пластины параллельно продольному сечению квадратного волновода позволяет создать на его основе волноводный поляризатор. Однако такая конструкция отличается большими продольными габаритами и ограниченной рабочей полосой частот.

Для устранения этих недостатков на боковых сторонах диэлектрической пластины размещается продольный проводник [6]. Реализация подобных поляризаторов возможна только на основе отрезков волноводов круглого или квадратного сечения. Размещение пластины внутри прямоугольного волновода приведет к значительному затуханию распространяющейся в нем электромагнитной волны.

Возможно создание поляризатора на основе отрезка волновода круглого сечения с расположенными вдоль его продольной оси диэлектрической пластиной с согласующими четвертьволновыми отрезками [7]. Симметрично относительно продольной оси в пластине через специальные отверстия закрепляются проводящие штыри. Штыри располагают перпендикулярно плоскости пластины. Четвертьволновые отрезки обеспечивают хорошее согласование для волн, чья поляризация параллельна плоскости пластины, а проводящие штыри (путем выбора их длины) – для волн, поляризованных перпендикулярно плоскости пластины. Однако недостаток подробной конструкции заключается в точности расчетов геометрических параметров и сложности изготовления ее элементов – незначительные отклонения от расчетных значений приводят к рассогласованию устройства.

Таким образом, можно выделить две основные сложности при проектировании и изготовлении волноводных поляризаторов: ограничение рабочих диапазонов частот и сложность конструктивной реализации, требующая повышенной точности в изготовлении, монтажных и крепежных работ. Поэтому, наравне с поиском технических решений для обеспечения улучшенных характеристик волноводных поляризаторов необходимо также применять дополнительные меры для их конструктивного упрощения. Одним из возможных решений этой проблемы является поиск и разработка новых материалов и структур, обеспечивающих одновременно миниатюризацию, упрощенную конструкцию и эффективность устройств, созданных на их основе.

Модель волноводного поляризатора на основе клиновидной магнитной стенки

В данной работе предлагается модель волноводного поляризатора, выполненная на основе отрезка прямоугольного волновода с широкой клиновидной магнитной стенкой [8]. Идеальная магнитная стенка (PMC – Perfect Magnetic Conductor – идеальный магнитный проводник) олицетворяет собой такие граничные условия, при которых все тангенциальные

компоненты магнитного поля и нормальные компоненты электрического поля равны нулю. Хотя такая модель является лишь математической абстракцией, на практике возможно создание структур, на поверхности которых касательная составляющая напряженности магнитного поля обращается в нуль.

Такие композитные структуры представляют собой совокупность периодически расположенных металлических грибовидных элементов и получили название грибовидных метаматериалов. Каждый элемент представляет собой одиночный колебательный контур, размеры которого много меньше рабочей длины волны. Переходное отверстие («ножка» гриба) и проводящий элемент («шляпка» гриба) образуют эквивалентную погонную индуктивность. Эквивалентная погонная емкость сосредоточена в зазоре между соседними элементами.

Периодическая структура грибовидного типа помещена в диэлектрическую среду. В качестве материала для диэлектрической подложки выбран распространенный в микроволновой технике материал Taconic RF-60TC – стекловолокно с керамическим наполнителем на основе фторопласта. Диэлектрическая проницаемость материала $\epsilon=6,15$.

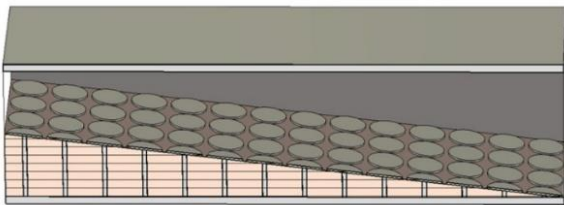


Рис.2. Волноводный поляризатор с клиновидной магнитной стенкой из грибовидного метаматериала

Магнитная стенка из грибовидного метаматериала выполнена в форме клина и располагается параллельно широкой стенке прямоугольного волновода типа WR-284 (размеры поперечного сечения $72,14 \times 34,04$ мм, рабочий диапазон от 2,60 до 3,95 ГГц) (см. рис.2). Геометрические размеры проводящих элементов идентичны друг другу. Длина переходных отверстий линейно увеличивается вдоль продольной оси волновода. Максимальная высота магнитной стенки равна половине высоты волновода. Так как ряды переходных отверстий равноудалены друг от друга в продольном сечении волноводной структуры, возрастание длины переходных отверстий пропорционально тангенсу угла наклона клиновидной магнитной стенки и расстоянию между рядами переходных отверстий. Расчеты проводились при высоте клиновидной магнитной стенки, равной 0,5 и 0,75 от длины узкой стенки (высоты) волновода H . Полученные результаты при высоте клина $0,5 \cdot H$ мм демонстрируют более выраженные поляризационные свойства, чем результаты при высоте клина $0,75 \cdot H$ мм.

Результаты компьютерного моделирования

Результаты компьютерного 3D-моделирования диаграмм направленности излучения из открытого конца волноводного поляризатора демонстрируют изменение характера поляризации распространяющейся электромагнитной волны. На частоте 600 МГц основные резонансные процессы в грибовидной структуре обусловлены емкостным характером метаматериала и наблюдаются только в тех рядах грибовидных элементов, чей диаметр проводящего элемента превышает линейные размеры переходного отверстия. В данном случае характер поляризации электромагнитной волны не меняется. Если возбуждать поляризатор волной с линейной горизонтальной поляризацией, излучение из открытого конца волновода будет также линейно поляризованным (см. рис.3).

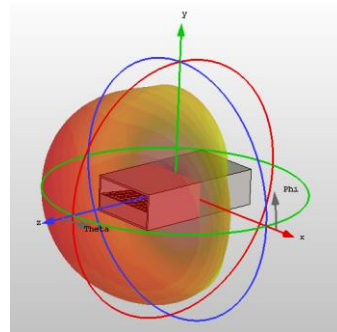


Рис.3. Диаграмма направленности на частоте 600 МГц

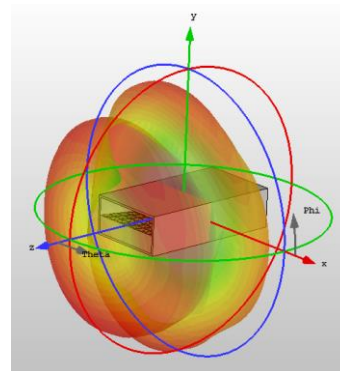


Рис.4. Диаграмма направленности на частоте 2 ГГц

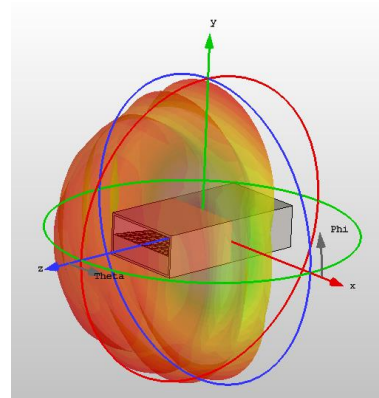


Рис.5. Диаграмма направленности на частоте 4 ГГц

По мере увеличения входной частоты до резонансной частоты грибовидного метаматериала характер поляризации излучения изменится с линейной горизонтальной на эллиптическую, что демонстрирует рис.4. На смену типа поляризации указывает форма диаграммы направленности излучения из открытого конца волноводного поляризатора. При этом резонансные процессы начинают сильнее проявляться с тех рядах грибовидных элементов, у которых диаметр проводящих элементов и высота переходных отверстий становятся приблизительно равными.

По мере дальнейшего увеличения частоты наиболее сильное проявление резонансов наблюдается в рядах, где высота переходных отверстий превышает диаметр проводящих элементов. Характер метаматериала становится индуктивным. Данное явление влечет за собой изменение поляризации излучения с эллиптической на линейную вертикальную, что изображено на рис.5.

Заключение

Представленные результаты свидетельствуют о возможности создания волноводных поляризаторов на грибовидных метаматериалах. Применение последних позволит расширить рабочий диапазон частот прямоугольного волновода, ограниченный критической частотой, а также обеспечить преобразование электромагнитной волны с линейной горизонтальной поляризацией в волны с другими видами поляризации. Кроме того, монтаж клиновидной структуры на широкой стенке волновода не требует дополнительных крепежных конструкций, что технологически упрощает устройство поляризатора. Плотное прилегание мета-структуры будет осуществляться за счет веса конструкции, что обеспечит минимизацию отражений.

Список литературы

1. Овчинникова Е.В. Излучатель на квадратном волноводе с пластинчатой замедляющей структурой и выступами / Е. В. Овчинникова, С. Г. Кондрачева, П. А. Шмачилин, Э. В. Гаджиев, А. В. Ширяева // Вопросы электромеханики. Труды ВНИИЭ. – 2022. – с. 33-40.
2. Патент № 2 265 258 Российская Федерация, МПК H01P 1/161 (2006.01). Волноводный поляризатор: № 2 265 258: заявл. 23.03.2004: опубл. 27.11.2005/ Беляев В.В., Богданов Ю.Н., Маюнов А.Т. – 7 с.: ил.
3. Mengyao Zheng and Jiaran Qi, "A Low-Cost High-Performance Dual-Band Diaphragm Circular Polarizer"// 2020 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting. – 2020.
4. Jin Tian, Geer Teni, Tongyu Ding, Jun Xiao and Qiubo Ye, "An Improved Dovetail Shape Circular Polarizer"// 2022 Cross Strait Radio Science & Wireless Technology Conference (CSRSWTC). – 2022.
5. Шаров Г.А. Волноводные устройства сантиметровых и миллиметровых волн. – М.: Горячая линия – Телеком. – 2022. – 640 с.: ил.
6. Патент № 2 037 921 Российская Федерация, МПК H01P 1/161 (2006.01). Волноводный поляриза-

тор: № 2 037 921: заявл. 29.05.1992: опубл. 19.06.1995/ Шалякин А.И. – 7 с.: ил.

7. Патент № 2 275 717 Российская Федерация, МПК H01P 1/161 (2006.01). Волноводный поляризатор: № 2 275 717: заявл. 18.10.2004: опубл. 27.04.2006/ Шалякин А.И. – 7 с.: ил.

8. Патент № 222 615 Российская Федерация, МПК H01P 1/165 (2006.01). Волноводный поляризатор на метаматериале: № 222 615: заявл. 02.06.2023: опубл. 12.01.2024/ Елизаров А.А., Машкова М.А. – 11 с.: ил.

УСТАНОВКА ДЛЯ ИЗМЕРЕНИЯ ПОКАЗАТЕЛЯ ПРЕЛОМЛЕНИЯ

*Посмитный С.В., Ботов М.А.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Данная работа посвящена разработке установки, решающей задачу определения показателя преломления (ПП) оптических материалов. Предлагаемый комплекс включает в себя универсальные устройства передачи и приема света, устанавливаемые на подвижный столик для определения угла поворота.

Введение

В настоящее время определение ПП материалов является важной задачей для человека. Для многих веществ, которые мы используем в жизни, требуется знание значения этого параметра. Например, знание ПП позволяет правильно рассчитать фокусное расстояние линз и определить угловые характеристики призм, что важно для создания оптических приборов: линзы, очки, приборы для хирургических операций. В строительстве, электронике, космической технике, медицине и многих других сферах нам необходимо знать значение этого параметра [1].

Схема установки

Для измерения ПП [2] необходимо разработать установку, состоящую из трех основных частей: вращающийся столик (рис. 1, г) для измерения угла поворота (угла), устройство для излучения света, устройство для приема света. Условная схема установки изображена на рис.1.

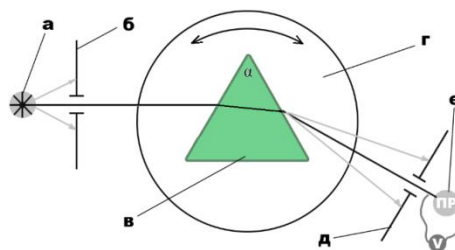


Рис.1. Схема установки для измерения показателя преломления

Устройство излучения света состоит также из нескольких частей: корпус устройства, источник питания, диод (рис. 1а), плата для смены диодов, резисторы, шелевая диафрагма для фокусировки (рис. 1, б). Корпус сконструирован таким образом, что имеется возможность изменять ширину щели для фокусировки с помощью двух регулируемых пластиковых пластинок.

Устройство для приема света также состоит из нескольких основных деталей: корпус устройства с регулируемым по ширине щелью (рис. 1д) фокусировки света, фотодетектор (рис. 1е), набор резисторов, источник питания, мультиметр.

Вращающийся столик содержит шкалу для измерения угла поворота. Устройства приема и передачи закрепляются на установке рядом со столиком. Основные части моделируются в программе AutoCAD и распечатываются с помощью возможностей 3D-принтера. Электронные компоненты установки соединяются между собой при помощи пайки и прикрепляются к распечатанным деталям корпуса. Готовые части закрепляются на неподвижном основании рядом с вращающимся столиком.

Процесс измерения показателя преломления

Устройство для излучения света подключается к источнику питания. На слот для замены диодов подается напряжение. В этот слот вставляется светодиод определенной длины волны. Пучок лучей проходит через диафрагму для фокусировки (рис.1б). После этого пучок проходит через трехгранную призму из исследуемого материала, преломляясь под определенным углом. Имея возможность поворота столика, мы подбираем максимальное значение напряжения на мультиметре – это будет означать, что прошедший призму луч, попадает в щель приемника (рис.1 д).

Для определения ПП n измеряют угол φ_1 , под которым луч попадает на призму с углом вершины α , расположенную на столике, а также угол отклонения луча ε . Углы измеряется при помощи засечек на вращающемся столике. Таким же образом определим угол отклонения луча. Эти параметры связаны друг с другом через соотношение [3]

$$\varepsilon = \varphi_1 - \alpha + \arcsin[n \sin\{\alpha - \arcsin(\sin\varphi_1/n)\}]. \quad (1)$$

Далее вычисляем значение n с помощью приведенной формулы (1), используя программу, разработанную на языке Python.

Заключение

В ходе проведенной работы была разработана установка, позволяющая определять ПП различных материалов. Полученные результаты эксперимента подтверждают эффективность работы установки.

Дальнейшее развитие данного исследования может включать расширение спектра изучаемых материалов, улучшение точности измерений, а также применение полученных данных для создания новых оптических устройств и технологий.

Список литературы

1. Юрин А. И., Вишняков Г. Н., Минаев В. Л. Измерение показателя преломления с помощью модифицированного метода постоянного отклонения // Измерительная техника. 2022. № 12. С. 35-39.
2. ГОСТ 28869-90 Материалы оптические. Методы измерений показателя преломления
3. Юрин А. И., Вишняков Г., Минаев В. Л. Измерение показателя преломления модифицированным методом призмы // Измерительная техника. 2023. № 2. С. 19-23.

РАЗРАБОТКА АВТОНОМНОГО ГАММА-СПЕКТРОМЕТРА: РЕШЕНИЕ ОСНОВНЫХ ЗАДАЧ

Сетдигов Р.М., Тюрин И.Г.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В данной работе исследуется один из вариантов создания автономного гамма-спектрометра на основе сцинтилляционного детектора, способного решать задачи в области измерения уровня радиации. В работе рассматриваются основные этапы проектирования электронного прибора, включая выбор элементной базы, разработку аппаратного обеспечения, а также тестирование и оценку его эффективности. Основное внимание уделяется обеспечению автономности работы прибора и его точности и надежности в изменяющихся условиях окружающей среды. Предложенные решения могут быть полезны для специалистов, занимающихся мониторингом радиационной обстановки, а также схемотехникой аналоговых приборов.

Введение

Во многих приложениях необходимо собирать данные с различных типов датчиков. Широкий спектр датчиков предоставляет полезную информацию в виде параметров импульсов: амплитуды, длительности импульса и других параметров. В данной статье рассматривается подход, направленный к измерениям, характерным для сигналов, несущих полезную информацию в виде пика (токовый импульс, поступающий от фотоумножителя), в котором амплитуда импульса пропорциональна энергии частицы. Для решения этой проблемы применяют следующие несколько способов, которые позволяют детектировать пики сигналов: аналоговый (с использованием диодов и фильтров) или цифровой с применением цифровой обработки сигналов. У популярного сцинтилляционного кристалла NaI(Tl) время свечения составляет сотни наносекунд, поэтому необходимо использовать аналоговое или аппаратное цифровое обнаружение максимального значения

импульса. В случае использования метода цифровой обработки сигнала для обеспечения высокого уровня разрешения необходимо использовать быстрые АЦП [1]. Для оценки требуемой скорости оцифровки необходимо оценить производную в окрестности пика, осциллограмма сигнала, полученного с датчика представлена на рис.1.

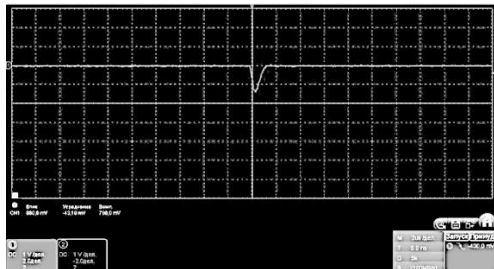


Рис.1. Осциллограмма сигнала сцинтилятора, зарегистрированная фотоумножителем с помощью используемого детектора

Недостатки метода максимальных значений и возможное их нивелирование

Для обработки сигналов с частотой порядка 1 МГц и более применяются микросхемы с программируемой логикой (ПЛИС). По сравнению с ПЛИС микроконтроллеры обладают некоторыми преимуществами такими как: универсальность, низкая стоимость, меньшее энергопотребление, а также представляют более простой процесс разработки приложений. Благодаря этим преимуществам микроконтроллеров в данной работе было принято решение разрабатывать устройство на их основе. АЦП недорогих микроконтроллеров общего назначения имеют ряд существенных недостатков, которые не позволяют их использовать в спектрометрии без дополнительной периферии. Именно поэтому в высокоточных стационарных приборах используют микросхемы ПЛИС, подключённые к внешним АЦП. Из основных недостатков – нелинейность АЦП микроконтроллеров и время выборки значений встроенной ячейкой выборки-хранения. Новизна предложенных решений состоит в методах, позволяющих устранить влияние этих факторов, которые описаны ниже.

Типовая частота работы АЦП микроконтроллеров – 2 МГц. Поэтому необходимо использовать схемы выделения максимального значения – вариации пиковых детекторов. Из теории обработки сигналов известно, что для получения корректной информации о форме сигнала требуется не менее 10 значений за период. Фронт исследуемого сигнала занимает порядка 200 нс, поэтому требуется оцифровывать с периодом в 40 нс, или с частотой не меньше 25 МГц. Из доступных на рынке микроконтроллеров – линейка PIC32 обладает наибольшей частотой оцифровки в 20 МГц, что недостаточно для разрабатываемого прибора.

При использовании пиковых детекторов – можно снизить требования с скорости АЦП до любого значения, удовлетворяющего требованию по загрузке датчика по условиям технического задания.

Один из методов оценки результатов при начальных этапах разработки – сравнить работу эталонного прибора с разрабатываемым прибором визуально (рис.2, рис.3).

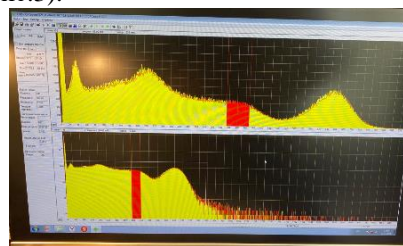


Рис.2. Спектр Cs-137, полученный с используемого детектора промышленным спектрометром Green Star SBS-75

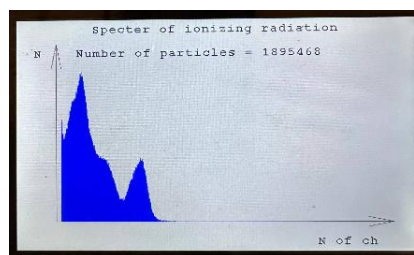


Рис.3. Спектр Cs-137, полученный с используемого детектора разработанным прототипом

При ограниченной скорости оцифровки в случае наложения двух импульсов предлагается диагностировать это наложение и игнорировать пришедшие в этот момент частицы. Диагностирование производится на основе длительности сигнала (рис.4).

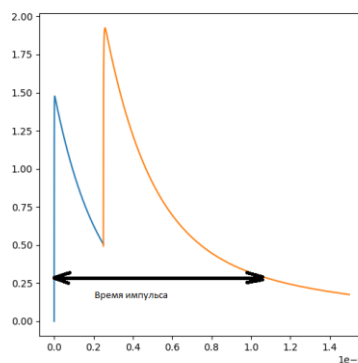


Рис.4. Изменение длительности импульса при наложении сигналов

Форма светового импульса зависит от температуры кристалла [2]. Поэтому необходимо это учитывать при проектировании алгоритма обработки.

Типовые проблемы и возможные варианты их решения

Одной из типовых проблем в области спектрометрии является температурная зависимость датчика [3]. Для переносного прибора, который применяется в различных условиях окружающей среды, этот фактор может критически отразиться на точности. У

фотоумножителя, который используется в качестве устройства, регистрирующего световые импульсы, нелинейно изменяется частота и амплитуда темновых импульсов (импульсы, возникающие в отсутствии светового сигнала), поэтому необходима схема, которая бы минимизировала темновые явления. Зависимости их частоты от температуры, полученные из документации [4], представлены на рис.5.

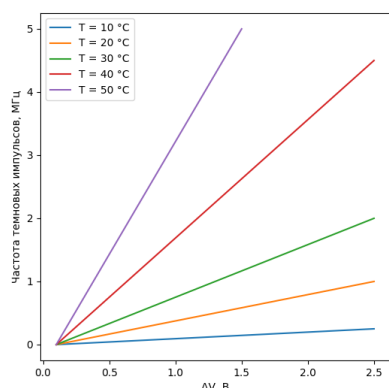


Рис.5. График зависимости частоты темновых импульсов от значения избыточного напряжения при разных значениях температуры

Так как темновые явления носят случайный характер, то в качестве решения предлагается использовать два фотоумножителя. Сравнивая два сигнала, можно точно сказать, ложным ли являлось срабатывание. Таким образом, необходима схема сравнения, которая будет пропускать только полезный аналоговый сигнал.

Принято считать, что аналоговые шумы описываются моделью гауссовского белого шума, поэтому было принято решение для увеличения соотношения сигнал-шум складывать сигналы с двух фотоумножителей. Этот подход позволяет выделить полезную информацию, хранящуюся в пике сигнала, с большей точностью.

Заключение

В ходе нашей работы были изучены тепловые зависимости датчика и осуществлены соответствующие доработки в алгоритм работы устройства. При разработке гамма-спектрометра основной упор был сделан на повышении энергоэффективности и осуществление автономности гамма-спектрометра. Были проанализированы различные варианты решений с целью оптимизации работы прибора. В результате был разработан гамма-спектрометр, обладающий требуемой точностью и надежностью в измерении уровня радиации. Развитие данного проекта может включать в себя дальнейшее усовершенствование в области энергосбережения и расширение функциональности прибора, что расширит область применения в различных отраслях науки и техники.

Список литературы

1. Cristian Z., Cătălin D., Cristian F. An FPGA Based Peak Detector for Magnetostrictive Current Sensors.

2. Saputra C., Waris A., Munir M. M., Novitrian (2022). Pulse Height Analyzer with Coincidence Correction. Journal of Physics: Conference Series, 2243(1). <https://doi.org/10.1088/1742-6596/2243/1/012037>

3. Ianakiev K. D., Alexandrov B. S., Littlewood P. B., Browne M. C. (2009). Temperature behavior of NaI(Tl) scintillation detectors. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 607(2), 432–438. <https://doi.org/10.1016/J.NIMA.2009.02.019>

4. What are the effects of temperature on dark count rates in an SiPM (MPPC)? Hamamatsu Photonics. (n.d.). Retrieved February 17, 2024, from <https://hub.hamamatsu.com/us/en/technicalnotes/mppc-sipms/what-are-the-effects-of-temperature-on-dark-count-rates-in-an-SiPM-MPPC.html>

МОДЕЛИРОВАНИЕ ФРАКТАЛЬНЫХ АНТЕНН КОХА И ПЕАНО С ДВУХСЛОЙНОЙ МЕТА-ПОВЕРХНОСТЬЮ

Воронова Е.И., Скрыбина В.А.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Отрицательный коэффициент преломления, наблюдающийся в ряде электродинамических систем, вызывает большой технический интерес и рассматривается во многих литературных источниках. Такие искусственные периодические структуры называются метаматериалами и имеют характеристики, не свойственные природным материалам. Микрополосковые антенны, покрытые слоем метаматериала, являются одной из наиболее интересных областей для изучения. В данной работе предложены фрактальные антенны трехслойной конструкции, содержащие один слой с геометрией Коха или Пеано, и два слоя метаматериала в виде периодических структур из модифицированных кольцевых резонаторов. Полученные результаты моделирования показывают улучшение коэффициентов усиления и направленного действия таких антенн.

Введение

Антенны являются незаменимой частью для всех систем беспроводной связи. Они служат связующим звеном между устройствами и свободным пространством или наоборот. Однако антенны имеют ряд ограничений, таких как работа на одной частоте, ограниченные коэффициенты усиления и направленного действия, физические размеры и эффективность [1].

В настоящее время актуальны антенны, обладающие следующими свойствами: компактность конструкции, широко- и многодиапазонность. Однако

получение подобных характеристик и конструкций антенн является сложной задачей. Классические микрополосковые антенны на многослойных печатных платах, стали наиболее популярным типом антенн из-за своих небольших габаритов, лёгкого веса и низкой стоимости производства [2], но такие антенны имеют низкий коэффициент усиления и узкую полосу пропускания [3]. Применение различных технологий при конструировании антенн позволяет эффективно улучшать характеристики антенн и их согласование по фидерной линии, тем самым решая некоторые из вышеперечисленных проблем. Одной из таких возможностей для преодоления ограничений классических антенн стало использование фрактальной геометрии, которая не может быть описана привычной Евклидовой геометрией [4]. Такой подход позволяет разработать новый тип фрактальных антенн, которые могут работать в нескольких диапазонах, при этом уменьшая размер антенны и увеличивая диапазон полосы пропускания.

Еще один способ, позволяющий улучшить характеристики антенны – это использование необычных свойств метаматериалов при проектировании [5]. Подобные материалы обеспечивают электромагнитные характеристики, не встречаемые в природе. Особый интерес получили отрицательные коэффициенты диэлектрической и магнитной проницаемости метаматериалов, возникающие в определенных частотных диапазонах [6]. Метаматериалы, созданные путем размещения разомкнутых кольцевых резонаторов (SRR) в периодическом массиве, являются одними из самых популярных конструкций. Подобные мета-поверхности, размещенные рядом с антеннами в качестве покрытий, улучшают коэффициент усиления антенны.

Использование мета-поверхностей в конструкции фрактальных антенн является новой областью, которая еще не исследована. В данной работе представлены топологии модифицированных кольцевых резонаторов: помимо основной щели, на разомкнутую сторону кольца добавляются дополнительные щели, выступающие в качестве распределенных конденсаторов. Такие кольца используются в двух слоях перекрывающейся подложки над антенной фрактальной формы. Для моделирования S-параметров исследуемых трехмерных структур используется программа High Frequency System Simulator (HFSS) компании AnSoft.

Конструкция и топология модифицированных кольцевых резонаторов

Размер одного кольца: 4 мм в длину и ширину, толщина кольца – 0,4 мм, ширина средней щели – 0,4 мм, ширина дополнительных щелей – 0,2 мм. В работе были использованы структуры с одной основной щелью (SRR1), с тремя (SRR2) и пятью (SRR3) щелями как показано на рис.1.

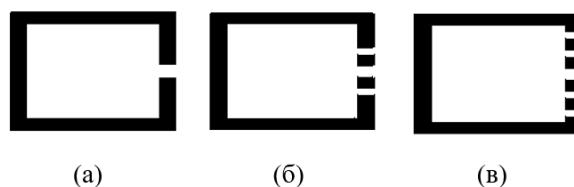


Рис.1. Модифицированный кольцевой резонатор: (а) – одна щель; (б) – три щели; (в) – пять щелей

Элементарные ячейки выполнены на FR4 подложке ($\epsilon_r = 4,4$) толщиной 0,25 мм. Расстояние между кольцами 1,4 мм. 225 колец равномерно распределены по подложкам, покрывающим антенны Пеано и Коха. Относительная диэлектрическая проницаемость, относительная магнитная проницаемость и коэффициент преломления определяются формулой Николсона-Росса-Вейра:

$$\epsilon_r = \frac{2}{jkd} * \frac{1 - v_1}{1 + v_1} \quad (1)$$

$$\mu_r = \frac{2}{jkd} * \frac{1 - v_2}{1 + v_2} \quad (2)$$

$$n = \pm \sqrt{\epsilon_r \mu_r} \quad (3)$$

где: $v_1 = S_{21} + S_{11}$, $v_2 = S_{21} - S_{11}$,

$k = \frac{\omega}{c}$, ω – угловая частота,

d – толщина подложки,

n – показатель преломления.

Конструкция и топология фрактальных антенн Коха и Пеано

1. Кривая Коха

Кривая Коха имеет бесконечную длину, поскольку на каждой итерации создается в четыре раза больше отрезков линий, чем на предыдущей итерации [7]. Таким образом, общая длина кривой увеличивается на одну треть с каждой итерацией. На рис.2 показано деление фрактальной структуры Коха для разных итераций. На нулевой итерации находится исходная длина линии. На следующих итерациях исходный отрезок делится на три участка, где средняя часть образует треугольную форму с длиной, равной одной трети исходной длины.

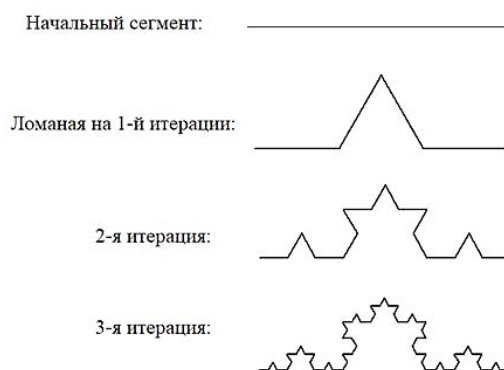


Рис.2. Три итерации кривой Коха

Исходная форма модели – это квадрат, длина стороны которого равна 20 мм. Генератор представляет собой вторую итерацию кривой Коха, примененной к каждой стороне квадрата. Результирующая форма модели представлена на рис.3. Микрополосковая антенна выполнена на подложке размером 65 x 65 x 1,5 мм. В качестве материала подложки используется диэлектрик FR4 с относительной диэлектрической проницаемостью $\epsilon_r = 4,4$.

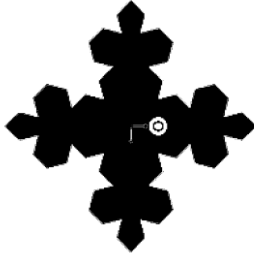


Рис.3. Форма фрактальной антенны Коха

2. Кривая Пеано

Рекурсивная процедура получения кривой Джузеппе Пеано представлена на рис.4. Генератор представляет собой прямую, поделенную на три части, средняя из которых также поделена на три. Средняя часть прямой сохраняется, а части, соседние с ней, удаляются. Далее средняя часть соединяется с оставшимися частями с помощью подстроечной длины, называемой шириной углубления.



Рис.4. Две итерации кривой Пеано

Изначально антенна представляет собой квадрат со стороной 28,5 мм. В качестве генератора выступает вторая итерация кривой Пеано. Генератор применяется к каждой стороне квадрата, образуя тем самым конечную форму фрактальной антенны, как представлено на рис.5. Микрополосковая антенна выполнена на подложке 60 мм x 60 мм x 1,6 мм. В качестве материала подложки используется FR4 с относительной диэлектрической проницаемостью $\epsilon_r = 4.4$

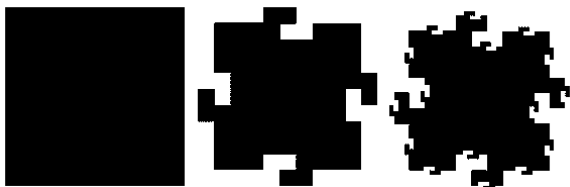


Рис.5. Форма фрактальной антенны Пеано

Результаты моделирования

Для проектирования и моделирования фрактальных антенн, покрытых двумя слоями метаматериала, было использовано программное обеспечение Ansoft HFSS, производящее моделирование методом конечных элементов.

1. Фрактальная антенна Коха

Расчеты S-параметров фрактальной антенны показывают, что эта модель обладает многодиапазонными свойствами. Она имеет резонансы на частотах 5,76 ГГц, 6,4 ГГц и 7,32 ГГц, что показано на рис.6. Диаграммы направленности излучения показаны на рис.7. Коэффициенты усиления и направленного действия на этих трех частотах приведены в таблице 1.

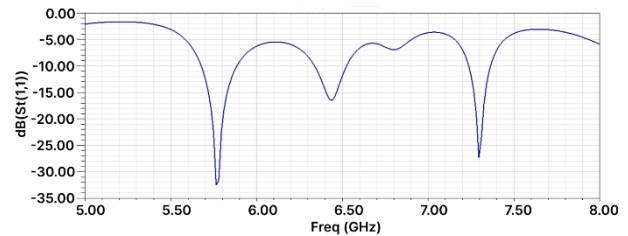


Рис.6. Резонансные частоты фрактальной антенны Коха без мета-поверхности

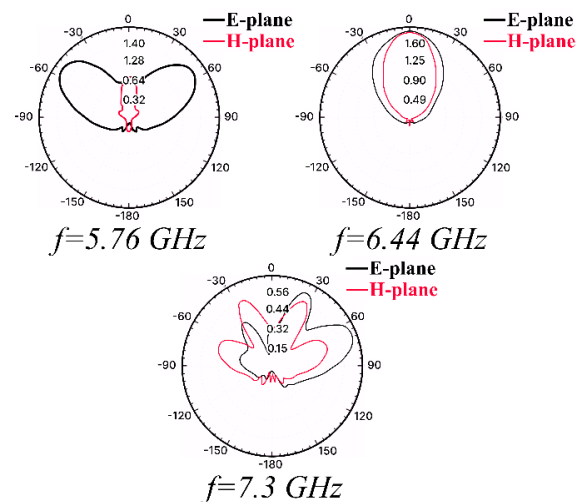


Рис.7. Диаграммы направленности фрактальной антенны Коха без мета-поверхностью (а) частота 5,76 ГГц (б) частота 6,44 ГГц (в) частота 7,3 ГГц

Таблица 1. Параметры фрактальной антенны Коха без мета-поверхности

| Frequency (GHz) | VSWR | Gain (dB) | Directivity (dB) | S11 (dB) |
|-----------------|------|-----------|------------------|----------|
| 5,76 | 1,06 | 6,1 | 8,9 | -32,4 |
| 6,4 | 1,12 | 7,7 | 10,4 | -16,4 |
| 7,29 | 1,08 | 2,1 | 8,1 | -27 |

Добавив к антенне два слоя мета-покрытия, мы получаем увеличение коэффициентов усиления и направленного действия. Также заметно улучшение согласования по фидерной линии на частотах 6,48 ГГц и 7,32 ГГц. На рис.8 и 9 показаны рассчитанные параметры коэффициента отражения S11 и формы диаграмм направленности излучения фрактальной антенны с мета-поверхностью соответственно. Результаты приведены в таблице 2.

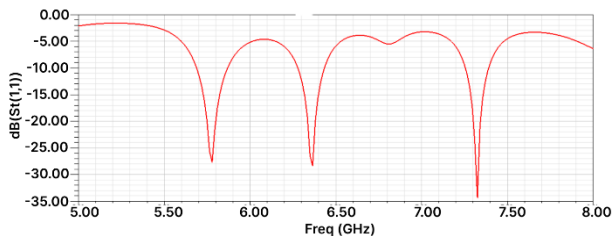


Рис. 8. Резонансные частоты фрактальной антенны Коха с мета-поверхностью

Таблица 2. Параметры фрактальной антенны Коха с мета-поверхностью

| Frequency (GHz) | VSWR | Gain (dB) | Directivity (dB) | S11 (dB) |
|-----------------|------|-----------|------------------|----------|
| 5,76 | 1,07 | 8,24 | 11,1 | -27 |
| 6,48 | 1,07 | 12,9 | 15,3 | -29 |
| 7,32 | 1,06 | 2,4 | 8,3 | -34 |

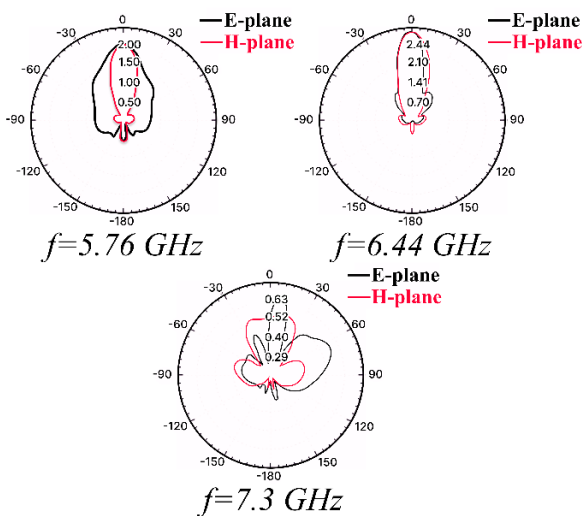


Рис. 9. Диаграммы направленности фрактальной антенны Коха с мета-поверхностью (а) частота 5,76 ГГц (б) частота 6,44 ГГц (в) частота 7,3 ГГц

2. Фрактальная антенна Пеано

Фрактальная антенна Пеано имеет резонансы на частотах 7,6 ГГц, 8,2 ГГц, 10 ГГц и 11,17 ГГц. Рис.10 отображает многодиапазонное поведение антенны с мета-поверхностью и без нее. Диаграммы направленности изображены на рис.11. Коэффициенты усиления и направленного действия антенны Пеано без мета-покрытия представлены в таблице 3.

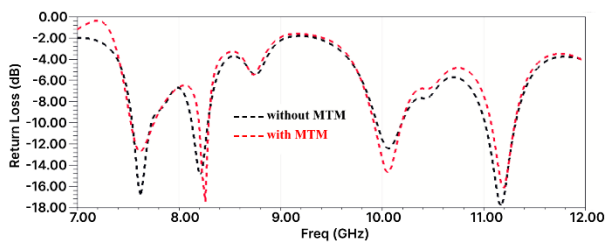


Рис. 10. Резонансные частоты фрактальной антенны Пеано с и без мета-поверхности

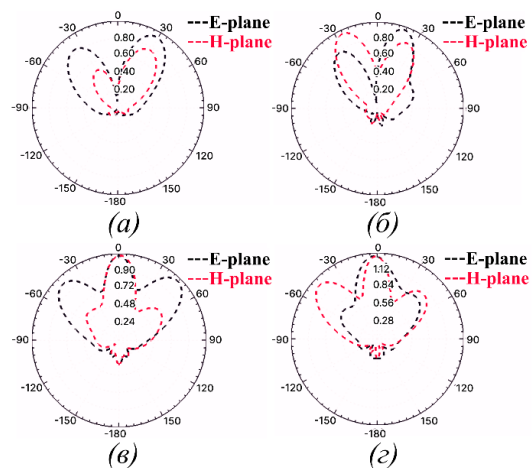


Рис. 11. Диаграммы направленности фрактальной антенны Пеано без мета-поверхности (а) частота 7,6 ГГц (б) частота 8,2 ГГц (в) частота 10 ГГц (г) частота 11,17 ГГц

Таблица 3. Параметры фрактальной антенны Пеано без мета-поверхностью

| Frequency (GHz) | VSWR | Gain (dB) | Directivity (dB) | S11 (dB) |
|-----------------|------|-----------|------------------|----------|
| 7,6 | 1,13 | 2,3 | 7,9 | -13,9 |
| 8,2 | 1,15 | 3,3 | 8,6 | -13 |
| 10 | 1,18 | 4,1 | 8 | -15,2 |
| 11,17 | 1,12 | 5,7 | 8,5 | -14,5 |

Добавление двух слоев мета-покрытия над антенной даёт улучшение коэффициентов усиления и коэффициентов направленного действия (КНД). Кроме того, наблюдается уменьшение КСВ на трех частотах (8,2 ГГц, 10 ГГц и 11,17 ГГц), что свидетельствует об улучшении согласования антенны. Рис.12 содержит диаграммы направленности фрактальной антенны с применением мета-поверхности. На частоте 8,2 ГГц коэффициент усиления достигает улучшения в 11,7 дБ. Значения улучшенных коэффициентов усиления и КНД представлены в таблице 4.

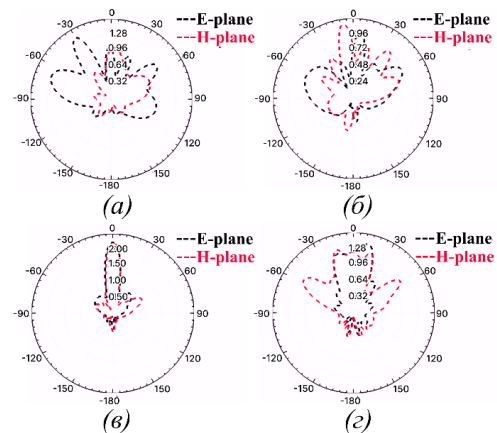


Рис. 12. Диаграммы направленности фрактальной антенны Пеано с мета-поверхностью (а) частота 7,6 ГГц (б) частота 8,2 ГГц (в) частота 10 ГГц (г) частота 11,17 ГГц

Таблица 4. Параметры фрактальной антенны без мета-поверхности

| Frequency (GHz) | VSWR | Gain (dB) | Directivity (dB) | S11 (dB) |
|-----------------|------|-----------|------------------|----------|
| 7,6 | 1,18 | 6,7 | 9,2 | -12,7 |
| 8,2 | 1,12 | 11,7 | 9,2 | -17,3 |
| 10 | 1,15 | 9,3 | 13,7 | -14,7 |
| 11,17 | 1,11 | 5,7 | 9,2 | -15,8 |

Заключение

В данной статье представлены результаты компьютерного моделирования фрактальных антенн с геометрией Пеано и Коха, в виде трехслойных конструкций, содержащих антенный слой и два слоя метаматериала в виде периодических структур модифицированных кольцевых резонаторов.

При этом для антенны Коха максимальное увеличение коэффициента усиления и направленности достигается при расстоянии 25 мм между самой антенной и первым слоем мета-поверхности, а расстояние между слоями составляет 4 мм.

Для антенны Пеано максимальный рост коэффициентов усиления и направленного действия достигается при расстоянии между антенной и первым слоем мета-поверхности, равным 15 мм. Слои метаматериала также находятся на расстоянии 15 мм друг от друга.

Следует отметить, что наличие мета-слоев, одновременно с улучшением параметров антенны, позволяет защитить её поверхность от негативных воздействий окружающей среды.

Список литературы

1. Сомов А.М., Старостин В.В., Кабетов Р.В. Антенно-фидерные устройства: Учебное пособие / Под ред. А. М. Сомова. — М.: Горячая линия Телеком, 2011. — 404 с.
2. Елизаров А.А., Закирова Э.А. Микрополосковые СВЧ устройства на печатных платах с многослойными диэлектрическими подложками. — М.: ИД Медиа Паблицер, 2016. — 122 с.: ил.
3. Елизаров А.А., Касторская А.С., Кухаренко А.С. Исследование микрополосковой антенны для радиочастотной идентификации // В кн.: Труды Российского научно-технического общества радиотехники, электроники и связи имени А.С. Попова. Выпуск LXV. Вып. 65. М.: РНТОРЭС им. А.С.Попова, 2010. С. 3-5.
4. Al-Saedi A., Perez J. Microstrip Fractal antenna geometries / Al-Saedi A., Perez J. // Przeglad Elektrotechniczny. — 2022. — № 8. — С. 98.
5. Kухаренко А.С., Шаймарданов Р., Елизаров А.А. A Methodology of Metamaterial Effective Permittivity and Permeability Value Measurement // Journal of Materials Science and Engineering A. 2017. No. 7-8. P. 206-215.
6. Елизаров, А.А., Кухаренко, А.С. Микрополосковые частотно-селективные устройства на резонансных отрезках электродинамических замедляющих систем и структурах с метаматериалами / А.А. Ели-

заров, А.С. Кухаренко; Нац. исслед. ун-т «Высшая школа экономики». — М.: Изд. дом Высшей школы экономики, 2019. — 68 с.

7. Hussin E., Kee J., Sahar N. Compact Koch Fractal Dipole RFID Antenna / Hussin E., Kee J., Sahar N. // Journal of Physics: Conference Series. — 2021.

РАЗРАБОТКА ГИПЕРСПЕКТРАЛЬНОГО ОПТИЧЕСКОГО МОДУЛЯ ДЛЯ БПЛА В ЗАДАЧАХ АТМОСФЕРНЫХ ИССЛЕДОВАНИЙ, ВАЛИДАЦИИ ДАННЫХ ДЗЗ И ОПТИЧЕСКОЙ СВЯЗИ В ОТКРЫТОМ ПРОСТРАНСТВЕ

Левашов С.Д.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В данной работе представлена концепция разработки и реализации модуля спектрометра для беспилотного летательного аппарата (БПЛА), который способен сканировать подстилающую поверхность, генерируя гиперспектральные данные в видимом и ближнем инфракрасном (ИК) диапазонах. Основная цель исследования заключается в разработке методики получения качественных гиперспектральных изображений, которые могут быть использованы для сравнительного анализа с данными дистанционного зондирования Земли (ДЗЗ), полученными со спутников, в целях улучшения алгоритмов атмосферной коррекции, а также для детального изучения возможностей дальней оптической связи.

Введение

Решение любых задач, связанных с прохождением излучения через атмосферу, будь то дистанционное зондирование земли (ДЗЗ) или оптическая связь, предполагает необходимость учета оптических характеристик атмосферы. В ситуациях базовой сложности расчеты могут быть выполнены с относительной простотой. Однако, в условиях умеренной сложности, когда детектирование целевого излучения еще возможно, но оценка влияющих факторов значительно усложняется, требуется использование специализированных инструментов для анализа получаемых данных или сигналов.

Таким инструментом является спектрометр, который мог бы сканировать подстилающую поверхность на небольшой высоте в умеренно сложных (для ДЗЗ) условиях. Это позволяет провести сравнение реальных данных со спутниковыми на выбранных спектральных диапазонах для валидации моделей атмосферной коррекции. Кроме того, БПЛА-спектрометр может быть использован для выполнения автоматических полетов на различные расстояния в условиях разной оптической прозрачности ат-

мосферы и там ожидать направленного лазерного сигнала. Это помогает в тестировании и исследовании разрабатываемых систем оптической связи, изучение их точности наведения и сбор статистики по шумам при различных мощностях излучения. В частности, использование двух лазерных каналов на разных диапазонах позволяет изучить относительные изменения их яркости.

Стоимость обычных беспилотников, оснащенных стандартными камерами для съемки в видимом диапазоне, значительно снизилась за последние годы, а надёжность и простота использования сильно улучшились. БПЛА с мультиспектральными камерами также становятся все доступнее. Это сильно расширяет круг исследователей из разных сфер, кому становится доступно дистанционное зондирование соответствующих масштабов. Появилось все больше профессиональных решений для получения гиперспектральных изображений, отличающихся от мультиспектральных большим количеством узких спектральных диапазонов [1]. Еще более простым и дешевым решением для получения данных с БПЛА в различных спектральных диапазонах может стать использование простого и легкого спектрографа, сконструированного таким образом, чтобы дополнять обычную стабилизированную камеру заводского дрона, не перегружая моторы стабилизатора. А основную работу по вычислению яркостей в спектральных диапазонах переложить на программное обеспечение.

Разработка технической части

Для описанных задач требуется БПЛА, способный подниматься на высоты порядка сотен метров, пролетать расстояние порядка 1-2 километров и уметь действовать в автоматическом режиме. Семейства Phantom и Mavic фирмы DJI вполне соответствуют всем этим требованиям и обладают достаточной надёжностью. Для первых опытов была выбрана уже устаревшая, но всё еще надёжная и недорогая модель Phantom 3 Advanced. В рамках проекта была разработана специализированная оптическая система на основе дифракционной решетки для генерации двумерной спектральной картины. Данный спектрометрический модуль был спроектирован таким образом, чтобы проецировать спектральное изображение непосредственно на объектив стандартной камеры БПЛА.

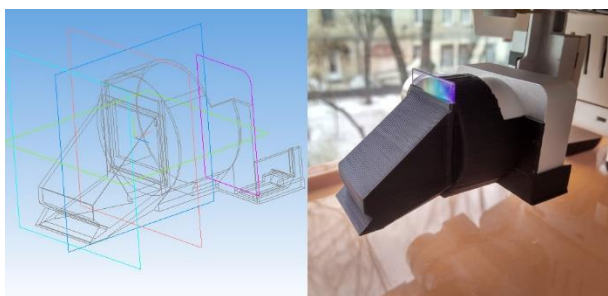


Рис.1. Модель и общий вид спектрального оптического модуля для БПЛА DJI Phantom 3 Adv

Конструкция корпуса модуля была разработана с использованием программного обеспечения Компас-3D для печати на FDM 3D-принтере, что позволило достичь точного соответствия с геометрией оригинальной камеры БПЛА, минимизировать массу изделия до 18 граммов без ущерба для прочности и защиты от проникновения света в ясную погоду. Такая легкость конструкции обеспечивает эффективную работу стабилизирующего подвеса камеры без риска перегрева двигателей. Стабилизация необходима для того, чтобы сохранить выбранное направление спектрометра в условиях постоянных наклонов и подрагиваний, свойственных мультироторным схемам БПЛА.



Рис.2. Общий вид. Полностью рабочий образец летающего спектрометра создан и проходит испытания. Спектральный модуль совмещен с камерой в передней части корпуса

В передней части этого модуля присутствует входное отверстие, выполненное в виде щели, что позволяет получить двумерную спектральную картину, формируемую видимой полосой на поверхности под дроном, зависящей от высоты полёта и угла наклона подвеса. При движении вперед спектральная картина меняется, происходит подобие сканирования push broom. Относительная простота технической реализации компенсируется современными возможностями компьютерной обработки полученных данных, а также методами калибровки, доступными на земле перед полетом и после.

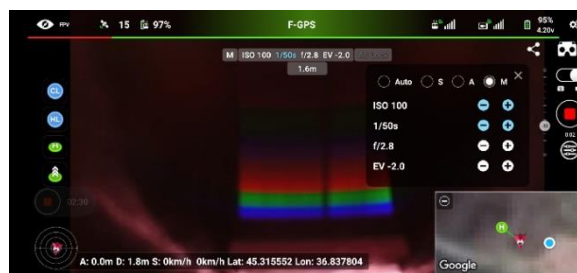


Рис.3. Спектральная картина, получаемая на стандартной камере и интерфейс программы автоматического полета по точкам

Обработка результатов

Для обработки полученных данных было написано ПО, позволяющее конвертировать снятую последовательность кадров в трехмерную матрицу, описывающую восходящее излучение в разных спектральных диапазонах. Сначала получившийся видеофайл разделяется на кадры, а с каждого кадра берется двумерная спектральная картина и нарезается на отдельные спектры, соответствующие каждый своей точке на полосе сканирования под БПЛА. Важно понимать, что сами по себе эти спектры не дают никакой важной информации и будут каждый раз отличаться, в зависимости от изменения условия освещения. Поэтому важно пересчитать эти спектры в доли от максимальной яркости, полученной при съемке равномерно освещенного белого листа. Таким образом мы получаем подобие Remote Sensing Reflectance (Rrs), т.е. отношение отраженного излучения к падающему. Результатом работы ПО является трехмерный массив яркостей, из которого можно вырезать любой слой для дальнейшей обработки. Или вырезать три слоя, соответствующие нужным спектральным каналам, назначить им цвета R, G, B и получить синтезированное псевдоцветное изображение.

Таким же образом можно выделить спектральные отрезки, соответствующие рабочим отрезкам действующих спутников ДЗЗ, провести сканирование в одно время с пролетом конкретного спутника и в случае удачных атмосферных условий провести сравнения полученных значений для одних и тех же полигонов на местности. Следующим этапом, основываясь на результатах сравнения, можно проводить подспутниковые измерения в более сложных атмосферных условиях и изучить расхождение используемых моделей атмосферы для спутников по данным уровней L0/L1 и уже валидированных, полученных с помощью БПЛА в самом низу атмосферы. Это позволит лучше изучить проблемы оптической связи на грани между хорошими атмосферными оптическими условиями и неприемлемыми.



Рис. 4. Маршрут БПЛА, границы полосы обзора спектрометра и спектральная картина, формируемая участком внутри белого прямоугольника

Также допустимо использовать такую модель летающего спектрометра в задачах разработки и наладки станций лазерной космической связи [2].

Как для оценки качества наведения и сопровождения тестовой цели (БПЛА), так и для изучения процессов, происходящих с оптическим каналом на существенных расстояниях. Это можно выполнять, фиксируя изменения яркости в соответствующих спектральных диапазонах одновременно для лазера наведения и телекоммуникационного лазера. Таким образом можно избежать установки двух детекторов со своими оптическими схемами и узкополосными фильтрами.

Полная обработка результатов гиперспектральной съемки должна включать в себя ряд корректирующих процессов, связанных в первую очередь с особенностью проецирования спектральной картины на камеру, а также с учетом использования фильтра Байера на матрице и как следствие – разной чувствительности элементов матрицы в разных участках спектра. Также требуется приведение всех полученных яркостей пикселей к координатам на местности и усреднение по сетке из ячеек (биннинг) требуемых размеров для возможности дальнейшего использования данных и/или их сравнения со спутниковыми изображениями. Точность соответствия длин волн на спектральном изображении будет зависеть от использования калибровочных источников излучения, либо совокупности фильтров с известными характеристиками, либо основываясь на данных о чувствительности используемой в камере матрицы.

Первые результаты тестов для спектрометра на базе камеры БПЛА Phantom 3 Advanced, производящей запись видео с разрешением 1920*1080 при 25fps, показывают что с разработанной оптической схемой при высоте полёта 30 метров, ширина полосы захвата на земле равна 30 метров и занимает 1/3 ширины матрицы, т.е. около 600 пикселей (крайние полосы из расчетов исключаются). Цифровое разрешение по продольной горизонтальной оси полностью зависит от скорости полёта и количества кадров при записи видео потока. Технические характеристики самого БПЛА позволяют подниматься до высоты 500м от точки взлёта (в пределах 5 тыс. метров), удаляться от пульта до 2-4км, держаться в воздухе до 23 минут.

Заключение

Было разработано и прошло начальные стадии тестирования устройство, которое может быть использовано в работе над спутниковой оптической связью, испытываемой между наземной станцией и малым космическим аппаратом формата Cubesat-6U, Импульс-1.

Список литературы

1. Adão T. et al. Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry //Remote Sensing. – 2017. – Т. 9. – №. 11. – С. 1110.
2. Reyes M. et al. Ground-to-space optical communication characterization //Free-Space Laser Communications V. – SPIE, 2005. – Т. 5892. – С. 589202.

СИСТЕМА СБОРА БИОПОТЕНЦИАЛОВ МЫШЦ У ЛЮДЕЙ С АМПУТАЦИЕЙ ВЕРХНЕЙ КОНЕЧНОСТИ ДЛЯ ВОСПРОИЗВЕДЕНИЯ МЕЛКОЙ МОТОРИКИ

Цуркан А.А.¹, Антонов М.А.²,
Коробок М.А.¹, Дуплин Г.В.²

Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент электронной инженерии,

²департамент прикладной математики

МИЭМ НИУ ВШЭ

Аннотация

В работе представлена система сбора данных, использующая данные электромиографии с ампутированной конечности и координаты положения здоровой руки пользователя, для воссоздания мелкой моторики на протезе.

Введение

Потеря конечности оказывает значительное физическое и психологическое воздействие на человека. Более того, люди с врожденным отсутствием верхних конечностей или те, кто перенес ампутацию, сталкиваются с значительными трудностями в повседневных задачах. Ампутация также влечет за собой серьезные экономические последствия, так как лишь небольшая доля людей с ампутацией может сохранить свою профессиональную деятельность. Современные протезы способны лишь частично решить эту проблему, так как только 27-56% людей с ампутацией заинтересованы в использовании протеза больше года [1].

Существуют два наиболее распространенных типа бионических протезов: механические и миоэлектрические – управляемые с помощью сигналов электромиографии [2]. Помимо бионических, многие используют косметические протезы, которые только создают видимость руки. Механические протезы, в свою очередь, довольно популярны. Они дешевле, чем более функциональные миоэлектрические аналоги, но обладают только базовыми возможностями, которые не делают их полноценной заменой руки [3].

Миоэлектрические протезы состоят из 2-32 поверхностных датчиков электромиографии (пЭМГ электроды), подключенных к основной системе управления [4]. Этот подход включает в себя использование сигналов активности мышц при их сокращении. Более продвинутые системы управления используют степень напряжения мышц при их сокращении для регулирования силы сжатия и скорости доведения жестов [5, 6]. Самой популярной системой управления в современных миоэлектрических протезах является система, реализованная через сокращение двух групп мышц: сгибающие и разгибающие мышцы предплечья. Это требует больше сознательных усилий при управлении протезом [7]. Используя комбинации сокращений этих мышечных групп, реализуется переключение до 41 предварительно записанных жестов в протезе [8]. Сокращения двух групп мышц предплечья, является функциональным под-

ходом, однако не наиболее оптимальным из всех. Его недостаточная интуитивность требует обширной подготовки, необходимой для использования системы. Кроме того, в данной системе управления есть вероятность ложного срабатывания во время работы протеза.

Улучшение управления возможно благодаря использованию сигналов от всех оставшихся мышц. Тем не менее, для обучения системы управления требуются точные движения пальцев, которые невозможно собрать на людях с ампутацией в силу отсутствия руки. Это приводит к значительной разнице в качестве распознавания движений руки с использованием сигналов пЭМГ у здоровых людей и потенциальными результатами для ампутантов. Например, авторы исследования сообщили о модели для декодирования движений пальцев в реальном времени из сигналов пЭМГ с медианным ошибочным углом в 6,24 градуса у здоровых людей [9].

Среда виртуальной реальности

Для создания целевых значений положений пальцев, которые соответствуют собранным биопотенциалам с ампутированной конечности, была использована технология виртуальной реальности. Данный подход позволяет реализовать симуляцию движений пальцев, используя зеркальное отображение кисти здоровой конечности. Метод зеркального отображения кисти здоровой конечности активно применяется в медицине и реабилитации [10]. Интеграция виртуальной реальности с зеркальной терапией представляет собой инновационный подход в лечении, предлагая значительные преимущества для восстановления [11]. Такой метод позволяет создать визуализированное, интерактивное пространство, в котором пациенты могут взаимодействовать со своей утраченной конечностью. Данный подход не только способствует улучшению нейропластичности [12], облегчая адаптацию мозга к изменениям, но и обеспечивает эффективную психологическую поддержку, ускоряя процесс адаптации к потере.

Для реализации среды виртуальной реальности были выбран очки “Oculus Quest 2” с поддержкой функции распознавания жестов. Данная функция позволяет с высокой точностью отслеживать движения пальцев и переводить их в движения виртуального объекта кисти. Сама виртуальная среда представляет собой комнату с экраном для демонстрации видео с жестами и элементами управления для переключения между ними.

Схема записи и агрегации данных о положении пальцев и активности мышц представлена на рис.1. С помощью очков виртуальной реальности здоровая рука визуализируется в среде соответственно положению. После модель руки зеркально отражается на место ампутированной. В это же время записываются данные активности мышц культи, которые синхронизируются с данными о положении отраженной кисти и сохраняются для дальнейшего использования.

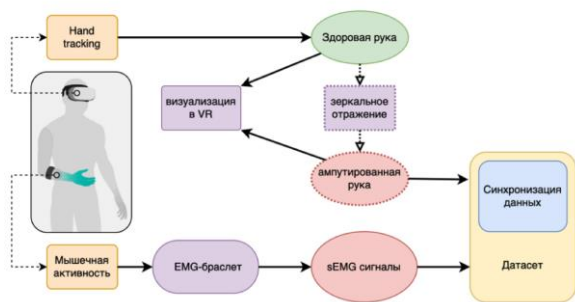


Рис.1. Логическая схема записи данных

Проектирование устройства регистрации биопотенциалов

Для определения характеристик биопотенциалов были проанализированы работы [13-17], в результате чего выдвинуты требуемые характеристики устройства:

- входной импеданс не менее 10 МОм;
- диапазон входного сигнала 0-10 мВ;
- усиление в пределах от 1000 до 50000;
- полоса пропускания 20-500 Гц;

Исходя из требований к устройству была предложена структурная схема (рис.2).



Рис.2. Структурная схема устройства регистрации биопотенциалов

Для реализации устройства регистрации биопотенциалов разработаны два вида печатных плат, одна из которых принимает пЭМГ-данные пользователя (рис.3), вторая оцифровывает и отправляет данные (рис.4). Печатные платы соединяются между собой шинами образуя браслет.

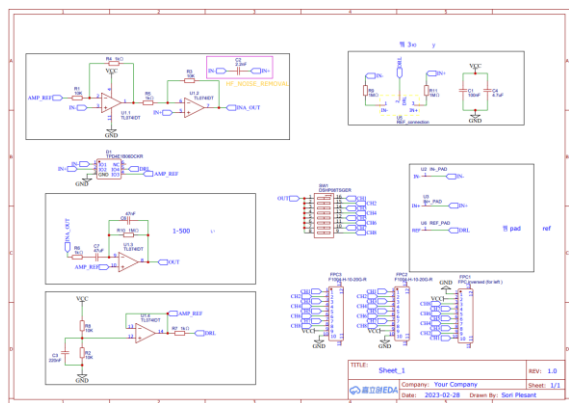


Рис.3. Принципиальная электрическая схема звена снятия пЭМГ-сигнала

В качестве основной микросхемы звена снятия ЭМГ-сигнала используется операционный усилитель TL074. На нем реализовано предусиление дифференциального сигнала, фильтр и задание точки нуля

для корректного считывания данных с помощью АЦП.

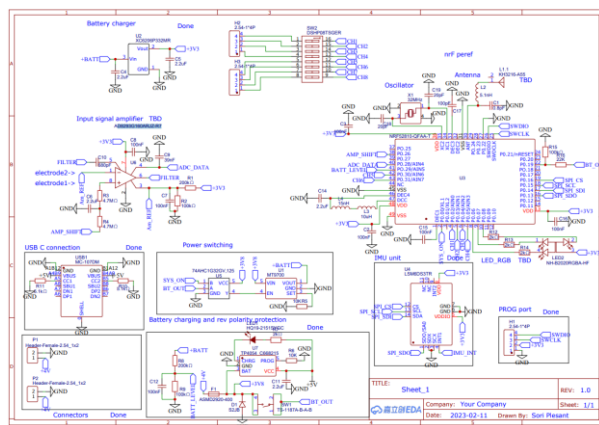


Рис.4. Принципиальная электрическая схема звена оцифровки и отправки пЭМГ-сигнала

Звено оцифровки и отправки пЭМГ-сигнала принимает данные со звеньев считывания сигнала и отправляет их на компьютер. Плата управляется контроллером NRF52810 поскольку данный контроллер имеет встроенный Bluetooth.

На рис.5 представлен экспериментальный образец устройства регистрации биопотенциалов.

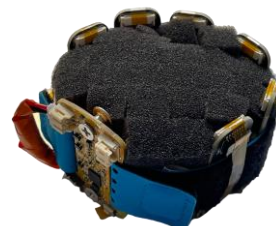


Рис.5. Экспериментальный образец устройства регистрации биопотенциалов

В результате было проведено исследование пЭМГ-активности с помощью экспериментального образца устройства. Записи данных различных движений представлены на рис.6-8.

По графикам видно, что в состоянии покоя сигнал состоит из шума малой амплитуды. При выполнении различных движений на графиках видна повторяемость сигнала по амплитуде, длительности и частоте.

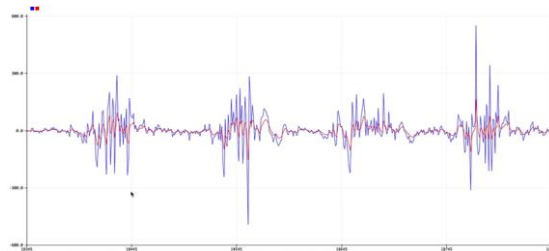


Рис.6. Запись данных при напряжении кулака

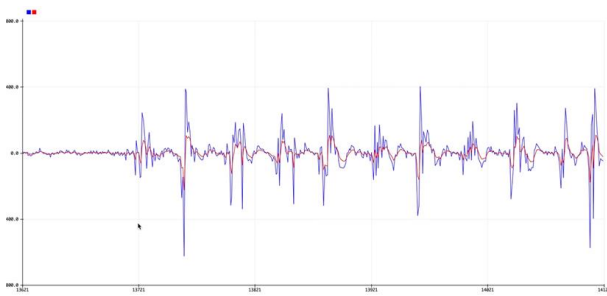


Рис.7. Запись данных при напряжении среднего пальца

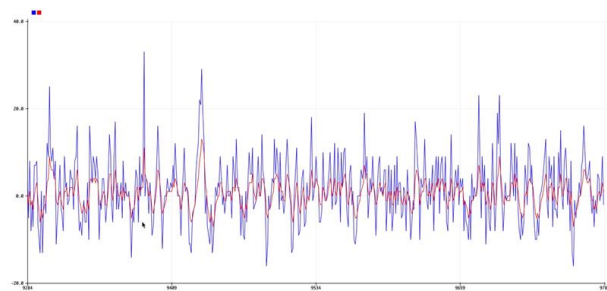


Рис.8. Запись данных в состоянии покоя

Заключение

В результате исследования выявлены ключевые характеристики биопотенциалов и представлена структурная схема их регистрации. Рассмотренные решения обеспечивают многоуровневое усиление и защиту от возможных помех, что значительно повышает точность получаемых данных.

Основываясь на данных, полученных в ходе исследования и повторяемости, полученной экспериментально, становится возможным объединение виртуальной реальности и браслета для считывания данных и обеспечение комплексного подхода для восстановления мелкой моторики у людей с ампутацией конечности и разработке протеза руки, включающей в себя данную систему. Рассматривая проблему и результаты шире, становится возможным создание универсального подхода к моделированию мелкой моторики через погружение в виртуальную реальность.

Список литературы

1. Raichle K.A., Hanley M.A., Molton I., Kadel N.J., Campbell K., Phelps E., Ehde D., Smith D.G. Prosthesis use in persons with lower- and upper-limb amputation // *J. Rehabil Res Dev* – 2008 – P.961-72.
2. Murray, C. D. Being like everybody else: the personal meanings of being a prosthesis user // *Disability and Rehabilitation* – 2009 – P.573-581.
3. Nagaraja, V. H., da Ponte Lopes, J., and Bergmann, J. H. M. Reimagining prosthetic control: A novel body-powered prosthetic system for simultaneous control and actuation // *Prosthesis* – 2022 – Vol. 4 – P.394-413.
4. Marinelli, A., Boccardo, N., Tessari, F., Domenico, D. D., Caserta, G., Canepa, M. Active upper limb prostheses: a review on current state and upcoming

breakthroughs // *Progress in Biomedical Engineering* – 2023.

5. Igual, C., Pardo, L. A., Hahne, J. M., and Igual, J. Myoelectric control for upper limb prostheses // *Electronics* 8 – 2019.

6. Li, J., Zhu, Z., Boyd, W. J., Martinez-Luna, C., Dai, C., Wang, H., et al. Virtual regression-based myoelectric hand-wrist prosthesis control and electrode site selection using no force feedback // *Biomedical Signal Processing and Control* 82 – 2023.

7. Zheng, Y., Chan, M., Shi, J., Chen, X., and Huang, Q. Sonomyography: Monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis // *Medical Engineering and Physics* 28 – 2006 – P.405-415.

8. Pizzolato, S., Tagliapietra, L., Cognolato, M., Reggiani, M., Müller, H., and Atzori, M. // Comparison of six electromyography acquisition setups on hand movement classification tasks. *PLoS ONE* 12 – 2017 – P.1-17.

9. Liu, Y., Zhang, S., and Gowda, M. NeuroPose: 3D Hand Pose Tracking using EMG Wearables // *Proceedings of the Web Conference* – 2021 – P.1471-1482.

10. Daniëlle, E, Raoul, M Bongers, Michiel, J. A. Jannink. Systematic review of the effectiveness of mirror therapy in upper extremity function // *Disability and Rehabilitation* – 2009 – V.31. – P.2135-2149.

11. Chris, H., Nadin, M., Tobias, L., Holger, R., Christian, D. Feasibility and psychophysical effects of immersive virtual reality-based mirror therapy // *Journal of NeuroEngineering and Rehabilitation*. – 2022 – V.19. – P.17.

12. Мирютова Н.Ф., Самойлова И.М., Минченко Н.Н., Цехмейструк Е.А. Терапевтические эффекты зеркальной терапии у больных после инсульта // *Вопросы курортологии, физиотерапии и лечебной физической культуры*. – 2021 – V.98. – P.14-23.

13. Меженная М. М. и др. Метод частотно-временного анализа суммарной электромиограммы в оценке функционального состояния нервно-мышечного аппарата человека // *Проблемы физики, математики и техники*. – 2012 – №. 1 (10). – С. 105-112.

14. Адигамова Р. Ф., Васильева Т. В., Мухаметвалеева Д. Р. ЭЛЕКТРОМИОГРАФИЯ // *Биотехнические, медицинские и экологические системы и комплексы. Биомедсистемы-2016*. – 2016. – С. 83-86.

15. Пестриков П. П., Пестрикова Т. В. Методы измерения сигналов электромиографии // *Информационные технологии XXI века*. – 2018. – С. 263-266.

16. Уразбахтина К. Р., Морозова Е. С., Уразбахтина Ю. О. Актуальные проблемы проектирования ЭМГ-датчиков // *Молодежный вестник УГАТУ. Технические науки-2022*. – 2022 – №. 1 (26). – С. 84-89.

17. Петрунин И. В., Кадырова Г. М., Мирина Т. В. Проектирование ЭМГ-датчика для миотонической системы управления протезом // *Молодежный вестник УГАТУ. Технические науки-2018*. – 2018. – №. 2 (19). – С. 84-89.

ПОЗИЦИОНИРОВАНИЕ ЧЕЛОВЕКА В ПРОСТРАНСТВЕ МИКРОВОЛНОВЫМ ИЗЛУЧЕНИЕМ НА ОСНОВЕ МЕТОДА ДОПЛЕРА

Ильянова Д.С.¹, Кацнельсон А.И.²,
Колпаков М.М.²

Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент электронной инженерии,

²департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Статья рассматривает применение инверсного радиолокационного синтезирования апертуры и эффекта Доплера для детектирования движущихся объектов в пространстве. Система, созданная на основе перечисленных технологий, позволит не только определять геолокацию, но и обнаруживать движущиеся объекты в темноте, зонах плохой видимости, местах, где они скрыты непрозрачными препятствиями по типу стены и так далее. Также обсуждаются возможности использования технологии сотовой связи и сетей Wi-Fi.

Введение

В современном мире использование микроволнового излучения стало повседневным [1]. Оно активно применяется в сотовой инфраструктуре для передачи данных и обеспечения связи между абонентами, предоставляя быструю и надежную связь в любой точке мира, сети wi-fi, которая делает возможным доступ в интернет в любом месте, где имеется соединение, тем самым позволяя эффективно осуществлять обмен информацией, работать удаленно, взаимодействовать с другими лицами, а также получать доступ к ресурсам и контенту в сети, и прочие технологии.

Из-за широкого распространения алгоритмов, предполагающих в своей основе использование микроволнового излучения, становится возможным применение выделяемого побочного излучения, не несущего в себе полезной информации, для нужд человека. На этой базе появляются различные системы, где энергия используется не только для приема-передачи информации, но и для питания электроприборов [2], подключения устройств умного дома, отслеживания геопозиции и многого другого.

Определение местоположения по сотовой связи

Развитая инфраструктура сотовой связи позволила определять геолокацию при помощи сотовой связи. Услуги предоставляют множество операторов связи, например: МТС [3], Мегафон [4] и другие. Принцип работы данной технологии заключается в том, что мобильный оператор может проследить и зафиксировать перемещения абонента путем анализа данных о сигналах, которые передаются между его мобильным устройством и сотовыми вышками, находящимися в его окружении. Технология создала новое направление в геоаналитике [5] и позволила

анализировать плотность населения, динамику его перемещения и другие метрики. Пример карты населения Новосибирска показан на рис.1.

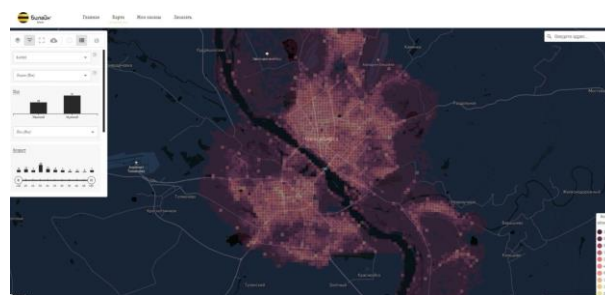


Рис.1. Карта численности населения на основе сотовых данных оператора Билайн

Сбор этих данных обеспечивает возможность определения приблизительного местоположения абонента с высокой точностью, достигающей нескольких метров.

Включив мобильный телефон или другое устройство в GSM-сети, запускается процесс поиска ближайших базовых станций. Телефон выбирает наилучшую соту, используя ее уникальный номер CellID и номер группы LAC. Затем Центр коммутации направляет вызов на нужную базовую станцию через контроллер BSC. Развитие геолокационных сервисов на базе GSM/UMTS-сетей было обусловлено тесной связью с отслеживанием местоположения в сотовой связи. Основные критерии работы таких сервисов включают точность и частоту определения координат абонента, а также возможность сохранения исторических данных и построения маршрутов передвижения. Для доступа к координатам абонента требуется его согласие. Различные сервисы, такие как Радар, Маячок и Навигатор, используют различные методы для точного определения местоположения пользователей. Сервис Контроль кадров позволяет дополнительно отслеживать вход и выход из геозоны, поиск сотрудников и агрегацию статистики маршрутов. Платформа геолокации обрабатывает данные с базовых станций и предоставляет информацию о координатах абонентов в реальном времени для различных сервисов, например, эта технология используется в экстренных службах для определения местоположения, звонящего по номеру 112.

Однако, для обеспечения работы такой системы необходимым условием является наличие у отслеживаемого объекта специального устройства, внешнего передатчика, при помощи которого осуществлялась бы связь. Вместе с тем, на смену описанным выше технологиям приходят такие устройства, для работы которых не требуется снабжать наблюдаемый объект дополнительным беспроводным устройством или набором датчиков.

Достичь такого результата возможно при помощи использования метода ISAR.

ISAR

Метод ISAR (Inverse Synthetic Aperture Radar) – это метод обработки радиолокационных данных, который применяется для изображения и анализа объектов на поверхности земли. Этот метод дает возможность получать качественные изображения объектов даже при ослаблении или искажении сигналов радиолокации. Радиолокационная станция посылает радиоволны к объекту, который движется относительно нее. Отраженные от объекта сигналы фиксируются станцией и служат для создания изображения объекта. После этого используются специальные алгоритмы обработки данных для устранения искажений, связанных с движением объекта, и для улучшения качества изображения.

С помощью этого метода можно отслеживать движущиеся объекты, не прибегая к использованию антенной решетки. Классическое устройство обнаружения может находить цель с помощью антенн [6], которые находятся на разных расстояниях друг от друга, и позже обрабатывать полученную информацию для определения направления, в котором расположена цель относительно решетки. В отличие от этого, в технологии ISAR используется только одна приемная антенна – объект сам является антенной решеткой, он действует как обратная синтезированная апертура, т.е. работает по такому принципу, что используется для формирования луча принимаемого сигнала во времени, а не в пространстве, и определения местоположения движущегося объекта, как следствие, в каждый момент времени делается только одно измерение. За счет того, что цель движется, последовательные измерения во времени создают эффект обратной антенной решетки, что похоже на то, как человек, двигаясь, показывает свое местоположение на устройстве. Обрабатывая такие последовательные измерения с использованием стандартного управления лучом антенной решетки, устройство может определить пространственное направление движения человека или животного (рис.2).

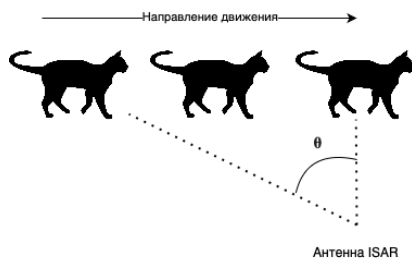


Рис.2. Пример работы ISAR

В классических системах отслеживание движущихся объектов осуществлялось за счет использования антенной решетки. Однако такие системы были дорогими и громоздкими, так как для осуществления корректной работы устройства требовалась массивная антенная решетка. Используя метод обнуления MIMO, устройство задействует преимущества антенной решетки, а затем применяет метод ISAR для избежания ее недостатков. ISAR использует только одну приемную антенну, поэтому она может фикси-

ровать только одно измерение в любой момент времени. Однако при перемещении объекта она делает выборку принятого сигнала в последовательных местоположениях в пространстве, словно у нее есть приемная антенна в каждой из этих точек. Благодаря взаимности каналов, последовательные временные выборки, полученные устройством, соответствуют последовательным пространственным местоположениям движущейся цели. Таким образом, устройство эффективно принимает во времени то, что антенная решетка принимала бы в пространстве. Используя последовательные временные выборки как пространственные выборки, устройство может эмулировать антенную решетку и использовать ее для отслеживания движения за стеной.

В этом контексте, эффект Доплера играет ключевую роль. Этот феномен, который был впервые описан австрийским физиком Кристианом Допплером в 1842 году, описывает изменение частоты волны для наблюдателя, движущегося относительно источника этой волны. Когда объект движется, частота принимаемого сигнала изменяется в соответствии с его скоростью и направлением. Это изменение частоты, известное как сдвиг Доплера, позволяет определить, в каком направлении движется объект. Это дает устройству возможность отслеживать движение объекта в реальном времени [7].

Предположим, что у нас есть выборка сигнала $y[\theta]$, которая получена от устройства в различные моменты времени θ и n . Эта выборка представляет собой угол отклонения между каналом, который соединяет устройство и отслеживаемого объекта. Значение θ положительно, если объект движется в том же направлении, что и устройство, и отрицательно в противоположном направлении. Теперь нужно определить меру сигнала вдоль пространственного направления θ в момент времени n , которую обозначим как функцию $A[\theta, n]$.

Для определения значения устройство обрабатывает полученную выборку, чтобы удалить эффект передаваемого сигнала и получить канал в виде $h[n] = y[n] / x[n]$. Для эмуляции антенной решетки размера w , она рассматривает w последовательных каналов, то есть $[n] \dots h[n + w]$. При применении стандартной антенной решетки

$$A[\theta, n] = \sum_{i=1}^w h[n + i] e^{j \cdot i \cdot \Delta \cdot \sin \theta \cdot \frac{2\pi}{\lambda}},$$

где λ – длина волны, а Δ – пространственное расстояние между последовательными антеннами в решетке.

Блок-схема устройства

Устройство представляет из себя 3 антенны [8], одна передающая и 2 принимающие (рис.3). Устройство состоит из HackRF One, как передатчика с рабочим диапазоном от 1 МГц до 6 ГГц, и 2 RTL2832SDR, как приемников, позволяя принимать диапазон с 60-1700 МГц.

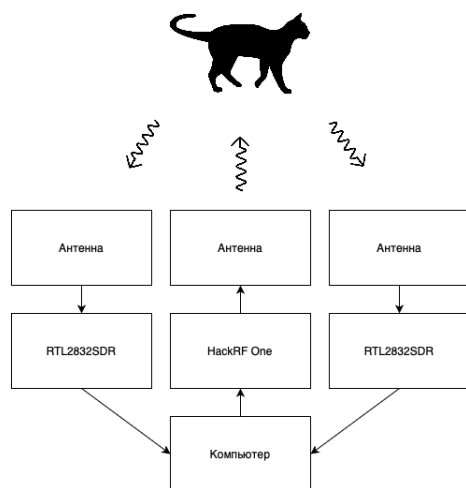


Рис.3. Блок-схема системы

Данная схема позволяет моделировать применение данной технологии к базовым станциям сотовых операторов, попадая в рабочий диапазон некоторых из них. Например, для 4G одной из рабочих частот для всех операторов связи в России будет 800 МГц.

Заключение

Проведенное исследование показывает возможность использования микроволнового излучения для обнаружения движения людей сквозь препятствия и внутри помещений. В ходе исследования была описана блок-схема устройства, которая послужит для дальнейшей реализации прототипа устройства. Данная технология предоставляет функциональность, которая ранее была доступна только в военных системах, позволяя обнаруживать движение с помощью доступных и экономичных устройств, работающих в общедоступных диапазонах. Уникальность данного подхода заключается в том, что для реализации данной системы не нужно создание новой инфраструктуры, а достаточно использование базовых станций сотовой или Wi-Fi связи. Применение технологии также может быть использовано экстренными службами для определения местоположения человека и животных.

Благодарность

Публикация подготовлена в ходе проведения исследования (проект №23-00-003 «Исследование технологий и устройств беспроводного трансфера электромагнитной энергии для высокоскоростных мобильных и нательных устройств Интернета вещей (IoT/IIoT) и киберфизических систем») в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2023–2024 гг.

Список литературы

1. Yelizarov, D. M. Ikonnikova, A. I. Katsnelson, P. I. Gromova and D. S. Ilyanova, "Monitoring of RF-Field Strength in the Environment by Mobile Telecommunications in the Moscow," 2023 Systems of Signals Generating and Processing in the Field of on Board Communications, IEEE, 2023, P. 1-4.

2. Yelizarov, D. M. Ikonnikova, O. V. Arinin, A. I. Katsnelson and D. S. Ilyanova, "Intelligent System for Wireless Transmission of RF-Energy for Mobile IoT Devices," 2023 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), IEEE, 2023, P. 1-5.

3. Мой Поиск - сервис для определения местоположения людей по номеру телефона [Электронный ресурс] // МТС Поиск. 2004. 26 марта. URL: <https://poisk.mts.ru/mts-poisk-authorise/> (дата обращения: 01.02.2024).

4. A. A. Yelizarov, A. D. Kuznetsov, I. V. Nazarov and I. R. Lavrukhin, "Prospects for Applications of Small-Sized Antennas on Flexible Substrates," 2020 Systems of Signals Generating and Processing in the Field of on Board Communications, IEEE, 2020, P.1-4

5. geo-analytics.in - сервис для анализа географических данных и определения расположения объектов [Электронный ресурс] // Geo Analytics. (б. д.). URL: <https://geo-analytics.in> (дата обращения: 01.02.2024).

6. A. Yelizarov, A. D. Kuznetsov, I. V. Nazarov and I. R. Lavrukhin, "Prospects for Applications of Small-Sized Antennas on Flexible Substrates," 2020 Systems of Signals Generating and Processing in the Field of on Board Communications, IEEE, 2020, P.1-4.

7. Adib, Fadel & Katabi, Dina. (2013). See Through Walls with Wi-Fi!. Computer Communication Review. 43. 75-86. 10.1145/2486001.2486039.

8. A. A. Skuridin, A. A. Yelizarov, I. V. Nazarov, E. A. Zakirova and D.M. Ikonnikova, "A Dual-Band Wi-Fi Rectifying Antenna for RF Energy Harvesting in Cyber Physical Systems", SYNCHROINFO 2021 Systems of Signal Synchronization Generating and Processing in Telecommunications 2021, pp. 1-4.

ИССЛЕДОВАНИЕ СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК КОНТАКТНЫХ РАДИОПОМЕХ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ОБЕСПЕЧЕНИЯ ЭЛЕКТРОМАГНИТНОЙ СОВМЕСТИМОСТИ НА МОБИЛЬНЫХ ОБЪЕКТАХ

Андреева А.А., Уткин Б.В.

Национальный исследовательский университет «Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В статье анализируются спектральные свойства контактных радиопомех для обеспечения электромагнитной совместимости на подвижных объектах. Представлены результаты разработки и методика оценки спектра контактных радиопомех для прогнозирования их характеристик при различных условиях. Разработанный алгоритм позволяет улучшить проектирование радиосвязи в условиях повышенной

электромагнитной активности на подвижных устройствах. Полученные рекомендации об использовании результатов прогнозирования помогают эффективно управлять уровнями контактных радиопомех для надлежащей работы радиооборудования на мобильных объектах.

Введение

Использование большого количества радиоэлектронных устройств на подвижных объектах в ограниченном пространстве приводит к возникновению контактных радиопомех, ослабляющих надёжность радиосвязи. Сложность обеспечения электромагнитной совместимости усугубляется высокой плотностью элементов конструкций радиопередатчиков и механическим воздействием, создающим перемещения элементов конструкций и модуляцию электромагнитного поля. Образование контактных радиопомех обусловлено переизлучением электромагнитного поля от токопроводящих конструкций с контактными соединениями [1]. Обнаружены недостатки в методах прогнозирования и анализа контактных радиопомех, а также неизученные причины подобных помех в радиосвязи на подвижных объектах.

Имеется ряд работ, посвященные вопросам образования контактных радиопомех, [1] [2] [3] [4] [5]. В данных работах рассматриваются общие положения теории образования контактных радиопомех, но в них отсутствуют достаточное обоснование методик прогнозирования контактных радиопомех и методов расчёта параметров контактных соединений, а также не исследованы детально причины возникновения контактных помех на мобильных объектах радиосвязи.

Электрическая и механическая модель контактного соединения

Для изучения формирования контактных радиопомех и спектральных характеристик приведем пример болтового контакта, изображенного на рис.1, как типичного механического соединения для металлических конструкций. Болтовое контактное соединение подвержено внешнему механическому воздействию, и под его влиянием изменяются электрические параметры контакта во времени. Механическая модель болтового контакта с двумя степенями свободы показана на рис. 2.

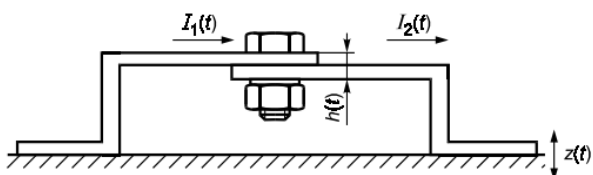


Рис.1. Болтовое контактное соединение под внешним механическим воздействием

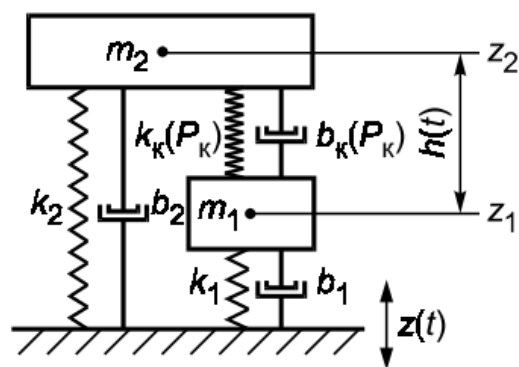


Рис.2. Эквивалентная механическая модель болтового контактного соединения

Модель можно математически описать с использованием системы дифференциальных уравнений [2], [3]:

$$\left. \begin{aligned} m_2 \ddot{z}_2(t) + b(P_K)(\dot{z}_2(t))^n + b_2 \dot{z}_2(t) - b_K(P_K)(\dot{z}_1(t))^n + \\ + (k_K(P_K) + k_2) z_2(t) - k_K(P_K) z_1(t) = k_2 z(t) + b_2 \dot{z}(t); \\ m_1 \ddot{z}_1(t) + b_K(P_K)(\dot{z}_1(t))^n + b_1 \dot{z}_1(t) - b_K(P_K)(\dot{z}_2(t))^n + \\ + (k_K(P_K) + k_1) z_1(t) - k_K(P_K) z_2(t) = k_1 z(t) + b_1 \dot{z}(t), \end{aligned} \right\} \quad (1)$$

где m_1, m_2 – эквивалентные сосредоточенные массы контактирующих деталей; $z_1(t), z_2(t)$ – эквивалентные сосредоточенные перемещения контактирующих деталей; $b_K(P_K), b_1, b_2, k_K(P_K), k_1, k_2$ – эквивалентные коэффициенты диссипации и жёсткости контакта и контактирующих деталей, $z(t)$ – внешнее кинематическое воздействие, n – степень зависимости диссипативной силы от скорости.

Система дифференциальных уравнений (1) представляет собой нелинейную систему, где параметры контактного соединения, такие как жесткость и коэффициент диссипации, зависят от внешних усилий и деформации. При нормальных механических нагрузках можно пренебречь указанным влиянием, что означает, что нелинейность будет влиять только на рабочую точку по указанным параметрам [7]. Нелинейность, вызванная степенной зависимостью диссипативной силы от скорости перемещения, может быть аппроксимирована линейной функцией из-за малых скоростей перемещения в области контактного соединения.

При высоких частотах протекающего тока изменения реактивных компонентов контактного сопротивления будут влиять на общее контактное сопротивление. Значительное значение динамической неустойчивости возникает в случаях, когда недостаточно сжатие болтового соединения.

Электрическая модель болтового контактного соединения представлена на рис.3.

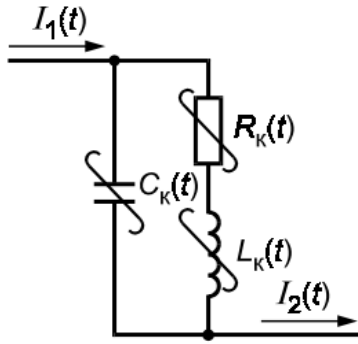


Рис.3. Электрическая модель болтового контактного соединения

Исследование, упомянутое в работе [6], демонстрирует, что под действием определенного механического воздействия в контактном соединении изменяются величины и фазы общего коэффициента передачи подвержены периодическим колебаниям. Это указывает на присутствие амплитудно-фазовой модуляции тока, протекающего через данное контактное соединение.

Полный коэффициент передачи по току для электрической модели болтового контактного соединения представляет собой сумму коэффициентов передачи по току в индуктивной $\dot{K}_L(\omega, R_c, L_c, C_c)$ и емкостной $\dot{K}_C(\omega, R_c, L_c, C_c)$ ветви:

$$\dot{K}(\omega, R_c, L_c, C_c) = \dot{K}_L(\omega, R_c, L_c, C_c) + \dot{K}_C(\omega, R_c, L_c, C_c) \quad (2)$$

Коэффициенты передачи по току в индуктивной и емкостной ветви определяются по формулам:

$$\dot{K}_L(\omega, R_c, L_c, C_c) = \frac{R_c^2 + (\omega L_c)^2 - \omega^2 R_c^2 C_c L_c - \omega^4 C_c L_c^3}{((1 - \omega^2 C_c L_c)^2 + \omega^2 R_c^2 C_c^2) \cdot (R_c^2 + (\omega L_c)^2)} + j \frac{(-\omega C_c R_c^3 - \omega^3 C_c L_c^2 R_c)}{((1 - \omega^2 C_c L_c)^2 + \omega^2 R_c^2 C_c^2) \cdot (R_c^2 + (\omega L_c)^2)} \quad (3)$$

$$\dot{K}_C(\omega, R_c, L_c, C_c) = \frac{\omega^4 C_c^2 L_c^2 + \omega^2 C_c^2 R_c^2 - \omega^2 L_c C_c + j\omega C_c R_c}{(1 - \omega^2 C_c L_c)^2 + \omega^2 R_c^2 C_c^2} \quad (4)$$

где ω – циклическая частота тока, проходящего через контактное соединение.

Определение вторичного тока и спектральных характеристик контактных радиопомех

Пусть через рассматриваемое контактное соединение будет проходить моногармонический ток, определяемый следующим выражением:

$$I_1(t) = I_{1m} \cos \omega t \quad (5)$$

При этом контактное соединение подвергается моногармоническому внешнему воздействию:

$$z(t) = z_m \cos \Omega t \quad (6)$$

Тогда выражение для вторичного тока можно получить путём разложения функций синуса и косинуса в полиномиальный ряд не выше третьей степени

вследствие того, что величина динамической неустойчивости мала:

$$I_2(t) = I_{1m} \left[\sqrt{\left(-\frac{Mm}{2}\right)^2 + \left(1 - \frac{M^2}{4}\right)^2} \cos(\omega t + \psi_1) + \sqrt{\left(\frac{m}{2} - \frac{3M^2 m}{16}\right)^2 + \left(\frac{M^3}{16} - \frac{M}{2}\right)^2} \left[\cos((\omega \pm \Omega)t + \psi_2) \right] + \sqrt{\left(-\frac{M^2}{8}\right)^2 + \left(-\frac{Mm}{4}\right)^2} \left[\cos((\omega \pm 2\Omega)t + \psi_3) \right] + \sqrt{\left(\frac{M^3}{48}\right)^2 + \left(-\frac{M^2 m}{16}\right)^2} \left[\cos((\omega \pm 3\Omega)t + \psi_4) \right] \right], \quad (7)$$

где m – амплитуда динамической неустойчивости модуля полного коэффициента передачи по току при частоте Ω внешнего механического воздействия; M – фаза динамической неустойчивости модуля полного коэффициента передачи по току; $\psi_1, \psi_2, \psi_3, \psi_4$ – начальные фазы.

Из анализа уравнения (4) следует, что спектр вторичного тока включает гармоники с частотами, кратными частоте внешнего механического воздействия. Следовательно, спектр вторичного электромагнитного поля шире, чем спектр изначального излучения электромагнитного поля.

При облучении электрического контактного соединения, которое находится в детерминированном динамическом режиме, наводимый ток будет претерпевает амплитудно-фазовые изменения [6]. Следовательно, спектр тока может определяться следующим выражением:

$$F = I_2 \left[\sum_{n=0}^{\infty} K_c \cos(\Omega t + \theta_n) \right] \cos \left(\omega_k t + \varphi_k + \sum_{n=0}^{\infty} \phi_{an} \cos(\Omega t + \theta_n) \right) \quad (8)$$

где I_2 – амплитуда тока вторичного излучения; K_c – модуль полного коэффициента передачи по току для болтового контактного соединения; φ_{an} – амплитуда фазы полного коэффициента передачи по току для болтового контактного соединения; ω_c – частота тока вторичного излучения; θ_n – начальная фаза полного коэффициента передачи по току для болтового контактного соединения.

Алгоритм прогнозирования контактных радиопомех

Беря во внимание полученные выражения, был разработан алгоритм прогнозирования контактных радиопомех.

1. Задаётся облучающее электромагнитное поле, вид и параметры контактного соединения.

2. На основе данных о контактном соединении выполняется построение его эквивалентной механической и электрической схемы.

3. Определяется аналитическое выражение для механической модели контактного соединения в виде системы дифференциальных уравнений и для электрической модели в виде частотного коэффициента передачи по току.

4. На основе построенной эквивалентной схемы контактного соединения определяется выражение для тока вторичного излучения.

5. Определяется аналитическое выражение для внешнего механического воздействия.

6. определяются спектральные характеристики контактных радиопомех вторичного излучения.

Заключение

Использование множества радиоэлектронных устройств на подвижных объектах создает контактные радиопомехи, затрудняющие радиосвязь. Высокая плотность элементов радиопередатчиков и механическое воздействие способствуют возникновению этих помех. Проблемы с электромагнитной совместимостью усугубляются. Эти помехи обусловлены переизлучением электромагнитного поля от токопроводящих конструкций. Наблюдаются недостатки в прогнозировании и анализе контактных радиопомех, а также неопределенность в причинах их возникновения на подвижных объектах в радиосвязи.

Список литературы

1. А. Ya. Klementenko, В. А. Panov, V. F. Sveshnikov, Contact interference to radio reception, Voenizdat, 1979, 116 p.
2. А. Р. Levin, Contacts of electrical connectors of radio electronic equipment, Sovetskoe radio, 1972, 216 p.
3. N. V. Zernov, V. G Karpov, Theory of radio-technical circuits, Energiya, 1972, 816 p.
4. R. Holm, Electric contacts : theory and application, Springer, Berlin, Heidelberg, 1967, 484 p.
5. N. V. Zernov, V. G Karpov, Theory of radio-technical circuits, Energiya, 1972, 816 p.
6. Грачев Н.Н., Лазарев Д.В. Влияние контактных помех при обеспечении электромагнитной совместимости на подвижных объектах радиосвязи // «Технологии электромагнитной совместимости»: Научно - технический журнал. - 2007, № 2 (21). - с. 22-32
7. Грачев Н.Н., Артамонов Д.О., Уткин Б.В., Development and analysis of models for the formation of contact radio interference under deterministic and random mechanical effects on the contact mechanical joints of moving objects//Proceedings of the 2021 IEEE International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS), IEEE, , 2021, С. 430-435.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ РАСЧЕТА И МЕТОДИКА ИЗМЕРЕНИЙ ПАРАМЕТРОВ АНИЗОТРОПНОГО МАГНИТОРЕЗИСТИВНОГО ДАТЧИКА УГЛОВОГО ПЕРЕМЕЩЕНИЯ

Денисов М.В.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе предлагается устройство стенда для измерений характеристик анизотропного магниторезистивного (АМР) углового датчика. Также приводится методика измерений для снятия характеристик: потребляемого тока, зависимостей выходных напряжений от угла поворота бесконтактного анизотропного резистивного датчика угла. Представлена математическая модель расчета вносимых погрешностей.

Введение

Современные автоматизированные электроприводы, устройства механизации и автоматизации обычно содержат выходное звено, значение угла поворота которого является сигналом обратной связи для управляющей системы.

В последние два десятилетия стали широко использоваться бесконтактные датчики угла на основе магниторезистивного эффекта или эффекта Холла, так как являются компактным решением, а выходные сигналы могут быть обработаны и оцифрованы. Кроме того, существуют решения, в которых интегральные схемы для аналоговой или цифровой обработки сигнала расположены в одном корпусе с чувствительным элементом.

В магниторезистивных сенсорах используется анизотропный магниторезистивный эффект (АМР-эффект), при котором наблюдается зависимость электрического сопротивления чувствительных элементов от направления воздействующего на них магнитного поля. Данный эффект наблюдается в полупроводниках и металлах [1]. В настоящее время для изготовления магниторезистивных датчиков наиболее часто используются тонкие плёнки пермаллоя или сплавов NiFeCo[2], в зависимости от необходимой чувствительности могут применяться различные топологии элементов схемы и материалы плёнок. При производстве подобных датчиков используются стандартные полупроводниковые технологии [3], что позволяет изготавливать датчики без существенных технологических сложностей.

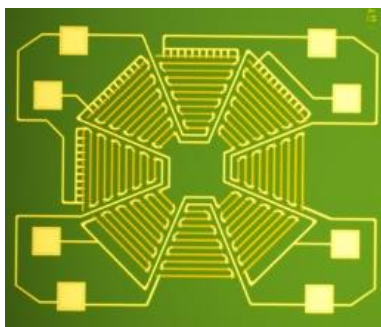


Рис.1. Топология АМР-сенсора

Разработкой и исследованием, а также улучшением алгоритмов обработки занимаются многие организации, специализирующиеся на выпуске датчиков для бытовых устройств (умные датчики, игровые джойстики и устройства ввода), датчиков для автомобильной и аэрокосмической промышленности (абсолютные энкодеры, датчики линейных и угловых перемещений).

Изучение научно-технической литературы, патентной документации, а также материалов конференций показало, что, несмотря на многолетнюю практику разработок и многомиллионное производство, вопросы обработки сигналов с последующей компенсацией погрешностей, вносимых различными факторами, изучены недостаточно.

В данной публикации будет рассматриваться датчик с однослойной топологией в виде двух измерительных мостов, показанная на рис.1, для которого характерна чувствительность к сильным магнитным полям: для корректной работы требуется напряженность поля не менее 25 кА/м, а магнитное поле земли не оказывает существенного влияния на выходные характеристики. Смещение двух измерительных мостов Уитстона выполнено в топологии таким образом, чтобы выходные сигналы имели синусную и косинусную зависимость от угла поворота напряженности магнитного поля относительно оси корпуса датчика.

Синусно-косинусный АМР-датчик

Датчик представляет собой прибор, выполненный в выводном корпусе SOIC-8. Во внутренних структурах датчика расположены 2 измерительных моста, смещенных относительно друг друга на угол 45 градусов. Все 4 сопротивления в мосте Уитстона в данном случае являются магниторезистивными элементами, что увеличивает чувствительность измерительной системы. Таким образом, достигается зависимость сигналов, которая может быть описана следующими выражениями:

$$\Delta U_1(\alpha) = U_{\text{пит}} S \cdot \sin(2\alpha) + U_{\text{см.1}} \quad (1)$$

$$\Delta U_2(\alpha) = U_{\text{пит}} S \cdot \cos(2\alpha) + U_{\text{см.2}} \quad (2)$$

где $U_{\text{см.1}}$ – напряжение смещение моста, обусловленное отличающимся исходным сопротивлением резисторов в мосту, S – коэффициент АМР-эффекта.

Таким образом, выходные сигналы приобретают зависимость от угла, показанную на рис.2.

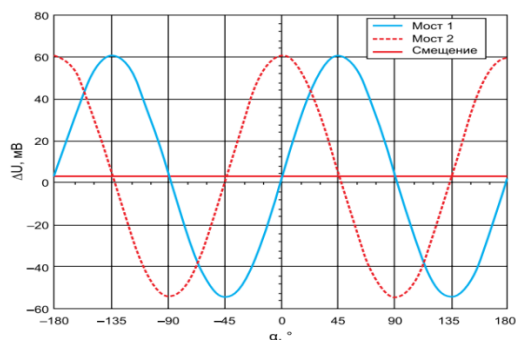


Рис.2. Зависимость напряжений на выходах от ориентации внешнего магнитного поля

Эффективное расстояние между корпусом прибора и магнитом может подбираться индивидуально, в зависимости от формы магнита и его магнитной силы, однако, для правильного функционирования требуется напряженность магнитного поля не менее 50 кА/м. Для диаметрально-намагниченного самарий-кобальтового магнита диаметром 5 мм и высотой 5 мм расстояние от корпуса должно быть в пределах 1,5 мм. По результатам моделирования в пакете ELCUT напряженность поля на расстоянии 1 мм до чувствительного элемента датчика составила 60 кА/м (см. рис.3), что подтверждается измерениями магнитометром.

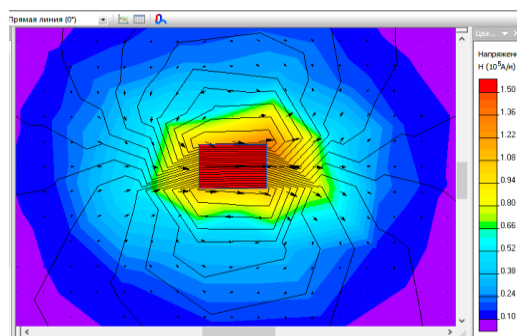


Рис.3. Моделирование магнитного поля диаметрально намагниченного цилиндрического магнита

Конструкция измерительного стенда

Для снятия характеристик чувствительного элемента углового датчика 1382НХ065 использовался измерительный стенд. Принципиальная схема представлена на рис.4.

Конструктивно данный измерительный стенд состоит из жесткого основания (1), с жестко закрепленной на нем механической делительной головкой (2) и подвижно закрепленного на направляющих платформы для исследуемых приборов (3). Печатная плата с исследуемым магниторезистивным датчиком (4) устанавливается на подвижную платформу (3) посредством винтового соединения. Подвижный магнит вращается шпинделем (5), установленный на коническую посадку в делительную головку. Делительная головка имеет шкалы угловых градусов (7) и минут (6). На минутной шкале располагается ручка для

поворота выходного вала. Точность установки составляет 1 минута.

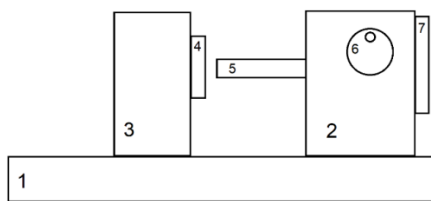


Рис.4. Конструктивная схема измерительного стенда

Схема измерений и результаты измерений

Для питания датчика использовался лабораторный источник постоянного напряжения PST-3202. Для измерения потребляемого тока и выходных напряжений использовался настольный мультиметр Arra-208. Схема измерений представлена на рис.5.

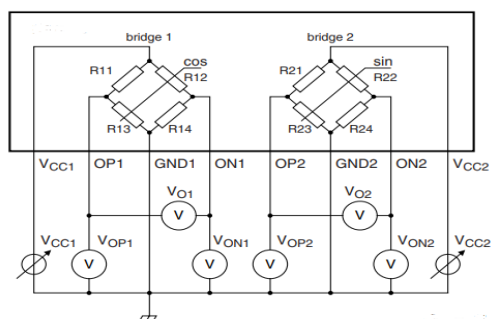


Рис.5. Схема измерений параметров мостовой схемы

По результатам измерений были получены и проанализированы данные, показывающие, что в полнооборотном режиме погрешность составила 0,4 угловых градуса [4], что показано на рис.6. Из иллюстрации видно, что искажения сигнала имеют периодический характер (8 периодов на оборот), кроме того, ошибка в основном отрицательная в одной половине окружности, и положительная – в другой.

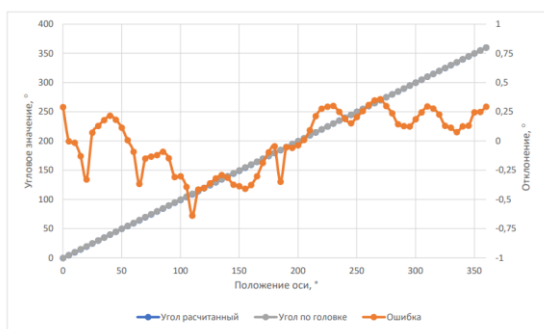


Рис.6. Отклонение вычисленного угла от действительного

Анализ отклонений

Наиболее простой частью отклонений можно считать температурную зависимость магнитосопротивления. Благодаря идеальной термической связи температурные коэффициенты обоих мостов практи-

чески одинаковы. магниторезистивный эффект имеет отрицательный температурный коэффициент. Это означает, что при более высоких температурах амплитуда сигнала уменьшается, а при низких – увеличивается.

Типичный температурный коэффициент (TCV_{peak}) для KMZ41 и его отечественного аналога 1382HX065 составляет -0,31 %/К. Следовательно, исходя из известного пикового напряжения при комнатной температуре, пиковое напряжение при любой другой температуре T может быть рассчитано следующим образом:

$$V_{peak_T} = V_{peak_{25^{\circ}C}}(1 + (T - 25^{\circ}C) * \frac{TCV_o}{100\%}) \quad (3)$$

Таким образом, зная температуру окружающей среды, можно внести корректировки в напряжение питания измерительного моста или в коэффициент усиления выходных сигналов.

На выходные характеристики так же оказывают влияние некоторые другие внешние факторы. Существуют три различные природы вносимых ошибок, которые могут быть вызваны неправильным расположением магнитного поля:

1) Отклонение от синусоидальной формы сигнала и гистерезис отклика датчика, при слабом магнитном поле, не насыщающем магниторезисторы (полное насыщение происходит только в бесконечно сильном магнитном поле).

2) Отклонения формы сигналов датчика, вызванные несимметричным расположением датчика относительно магнита (неоднородное магнитное поле).

3) Влияние внешних магнитных полей, влияющих на первичное поле, используемое для измерений. Влияние внешних полей невозможно описать в общих чертах, поскольку эти эффекты зависят от фактической схемы измерения. Единственная возможность избавиться от внешних полей или ограничить их воздействие – это использование магнитного экранирования.

Наибольшие неточности в измерения могут быть внесены особенностями конструкции измерительной системы, так как сложно установить магнит и датчик идеально соосно. Причина этих деформаций заключается в том, что чувствительная часть датчика не полностью помещается в однородную часть магнитного поля, или, другими словами, соответствующая часть магнитного поля, используемая для измерений, становится неоднородной. Поскольку фактическая угловая ошибка, возникающая из-за неоднородного поля, зависит от конкретной установки, ее невозможно рассчитать в целом. Однако для простого блочного магнитного устройства, упоминаемого в данной публикации, можно использовать эмпирическое правило[4]:

$$E_{неодн.} = C \cdot R^2 \quad (4)$$

где C – магнитная константа

$$C = 320^{\circ} / (w + l)^2 \quad (5)$$

где R – радиус, показывающий расстояние между центрами магнита и центром чувствительного элемента (см. рис.7), w – ширина магнита, l – длина магнита.

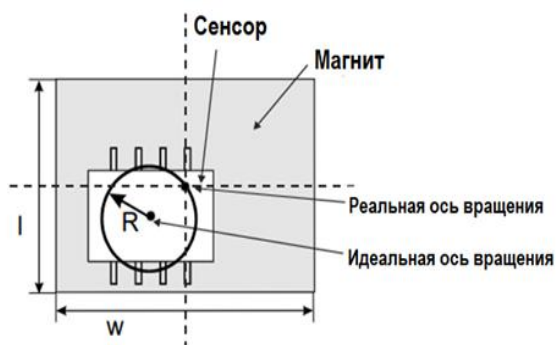


Рис.7. Несоосность магнита и чувствительного элемента датчика

В рассматриваемом случае можно сделать вывод, что магнитная константа $C \approx 3,2^\circ / \text{мм}^2$. Таким образом, несоосность в 1 мм может дать ошибку в 3,2 угловых градуса. Из этого можно сделать вывод, что при увеличении магнита или магнитного контура можно снизить требования к допускам позиционирования деталей в датчике.

Недостаточное магнитное поле приводит к двум эффектам. Первый из них — ошибка формы сигнала, вызванная несинусоидальной формой выходных сигналов. Из-за своей геометрической природы максимальные и минимальные значения всегда будут находиться в одних и тех же местах для каждого датчика. Максимальные значения наблюдаются при механических углах 11,25°, 33,75°, 56,25°, 78,75°, 101,25°, 123,75°, 146,25° и 168,25°. При этом ошибок измерения не возникает при 22,5°, 45°, 67,5°, 90°, 112,5°, 135°, 157,5° и 180°.

Синусоидальные выходные сигналы могут иметь смещения (постоянную составляющую), ограничивающие точность измерения. С точки зрения приложения смещение каждого канала можно разделить на две части: постоянная часть, которая практически устраняется компенсацией смещения и часть, меняющаяся с температурой.

$$X = X_0 \sin 2\alpha + \Delta x \quad (6)$$

$$Y = Y_0 \cos 2\alpha + \Delta y \quad (7)$$

где, Δx и Δy – отклонения по каждому из каналов. Из этого, ошибку, вызванную смещением можно описать следующим выражением:

$$E_{\text{offset}}(\alpha, \Delta x, \Delta y) = \left| \alpha - \frac{1}{2} \arctan \left(\frac{X_0 \sin 2\alpha + \Delta x}{Y_0 \cos 2\alpha + \Delta y} \right) \right| \quad (8)$$

Если принять, что $\Delta x = \Delta y$, то максимальная ошибка смещения можно выразить выражением:

$$E_{\text{offset_MAX}} = 0,4^\circ / \% \text{ Amp.} \quad (9)$$

Это показывает, что максимальная вносимая погрешность будет различной в зависимости от температуры и составляющих Δx и Δy , так как амплитуда зависит от температуры, что показывалось ранее. Так же из этого следует, что расположение области с наименьшими ошибками определить невозможно, так как на разных датчиках величины смещений и амплитуд сигналов при нормальных условиях будут отличаться.

Разница амплитуд сигналов косинусного и синусного измерительных мостов так же вносит вклад в формирование ошибки. Так как, несмотря на исполнение этих двух измерительных мостов на одной подложке, небольшие неоднородности материала могут повлиять на разность амплитуд сигналов. Вносимую ошибку можно выразить выражением:

$$E_{\text{Amp.}}(\alpha, A) = \left| \alpha - \frac{1}{2} \arctan \left(A \frac{\sin 2\alpha}{\cos 2\alpha} \right) \right| \quad (10)$$

где $A = \frac{X_0}{Y_0}$ – отношение амплитуд

При максимальном значении этого отношения, максимальная ошибка может быть выражена как:

$$E_{\text{Amp_MAX}} = 0,158^\circ / \% \quad (11)$$

Данная ошибка является периодической, с периодом 90 угловых градусов, и показывает, что вносимая ошибка будет равна 0,158° на каждый процент разницы между выходными сигналами.

Так же имеет место фазовая ошибка между каналами. Это означает, что фаза между сигналами X и Y не составляет точно 90° во всем измеряемом диапазоне. Исходя из этого, можно выразить выходные сигналы как:

$$X = X_0 \sin(2\alpha + \Delta\beta(\alpha)) \quad (12)$$

$$Y = Y_0 \cos 2\alpha + \Delta y \quad (13)$$

Так как сдвиг является функцией α , и не может быть постоянным на всём диапазоне измеряемых углов, угловая погрешность, вносимая разностью фаз, может быть выражена как:

$$E_{\text{Phase}}(\alpha, \Delta\beta) = \left| \alpha - \frac{1}{2} \arctan \left(\frac{\sin(2\alpha + \Delta\beta(\alpha))}{\cos 2\alpha} \right) \right| \quad (14)$$

Тогда, максимальная ошибка, вносимая фазовым сдвигом, будет определяться следующим выражением, где PS – фазовый сдвиг.

$$E_{\text{Phase_MAX}} = 0,5^\circ / \% PS \quad (15)$$

Исходя из выражений (4), (9), (11), (15) можно сделать вывод о необходимости точного монтажа компонентов в измерительной системе. При малом допуске несоосности (0,1 мм или менее), ошибкой, вызванной неоднородностью поля, можно пренебречь, так как наибольшее влияние будут иметь другие факторы, обусловленные особенностями изготовления датчиков. Тогда, результирующая ошибка выходного сигнала будет выражена как:

$$E_{Signal_MAX} = \sqrt{E_{Offset_MAX} + E_{Amp_MAX} + E_{Phase_MAX}} \quad (16)$$

Используя выражение (16) и измеренные значения, получаем, что для нашего случая максимальная ошибка сигнала будет составлять 0,46 угловых градусов.

Заключение

В данной работе предложена методика измерения выходных характеристик и выполнены эксперименты с анизотропным магниторезистивным датчиком углового перемещения. Представлена математическая модель расчета вносимых погрешностей, влияющих на точность преобразования магниторезистивного углового датчика.

Применение полученных в ходе исследования данных может помочь улучшить алгоритмы обработки измерительных характеристик вычислений.

Список литературы

1. Аш. Ж. и др. Датчики измерительных систем. Книга 1. – М. Мир, 1992 – 480 с.
2. Ажаева Л.А., Грабов А.Б., Суханов В.И. АМР-Сенсор индукции магнитного поля земли на основе тонких пленок сплава NiFeCo // Датчики и системы. – 2020. - №1. – с. 35-41.
3. Tumanski S. Thin film magnetoresistive sensors. – IOP Publishing Ltd, 2001. – 441 p.
4. Денисов М.В. Разработка интегрального датчика угла поворота на основе анизотропного магниторезистивного эффекта // Ежегодная межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов имени Е.В.Арменского. – 2021. - С. 192-194
5. APPLICATION NOTE Contactless Angle Measurement Using KMZ41 and UZZ9000 Philips Electronics N.V. 2000 52p.

РАЗРАБОТКА УСИЛИТЕЛЯ БИОЭЛЕКТРИЧЕСКИХ ПОТЕНЦИАЛОВ С БЫСТРЫМ МУЛЬТИПЛЕКСИРОВАНИЕМ ЭЛЕКТРОДОВ ДЛЯ ПОРТАТИВНОЙ СИСТЕМЫ РЕГИСТРАЦИИ ПОВЕРХНОСТНОЙ ЭЛЕКТРОМИОГРАММЫ ВЫСОКОЙ ПЛОТНОСТИ

Носов И.А.¹, Игнатьев Д.К.², Янцен Е.Э.²
Национальный исследовательский университет
«Высшая школа экономики»,
¹*департамент компьютерной инженерии,*
²*департамент прикладной математики*
МИЭМ НИУ ВШЭ

Аннотация

В данной работе исследуется проблема разработки компактного и энергоэффективного усилителя биопотенциалов для регистрации поверхностной электромиографии высокой плотности. Предложена архитектура с использованием быстрого мультиплексирования электродов. Разработано схемотехническое решение для усилителя биопотенциалов. Соотноше-

ние сигнал/шум (SNR) полученной системы – 18,7 дБ, среднеквадратичное значение шума – 19 мкВ.

Введение

Поверхностная электромиография высокой плотности (High-density surface electromyography, HD-sEMG) представляет собой метод, основанный на регистрации электрической активности мышц с высокой разрешающей способностью через кожный покров. В отличие от обычной поверхностной электромиографии, HD-sEMG позволяет получить данные об активности отдельных областей мышц за счёт использования большего числа электродов, расположенных близко друг к другу [1].

В последние годы активно исследуется возможность применения HD-sEMG в носимых биоинтерфейсах для решения задачи распознавания жестов кисти руки и восстановления сложных движений на основе данных о мышечной активности [2]. Такой подход позволяет отслеживать движения руки даже в случае отсутствия части конечности, что может использоваться в передовом протезировании [3].

Существующие коммерческие решения для регистрации HD-sEMG, используемые в исследованиях, обычно являются стационарными и имеют высокую стоимость. Такие решения требуют питания от сети и не могут использоваться в носимых устройствах [4].

Целью данной работы является разработка компактного усилителя биоэлектрических потенциалов с применением быстрого мультиплексирования электродов и низким энергопотреблением для использования в носимом биоинтерфейсе.

Стандартная архитектура для регистрации биопотенциалов

Сигналы электромиограммы характеризуются небольшой амплитудой, варьирующейся от нескольких микровольт до нескольких милливольт, и широким спектром частот от 5 Гц до 500 Гц, при этом самая важная информация содержится в полосе частот от 50 до 150 Гц [5]. Регистрация таких сигналов является сложной задачей не только из-за параметров самих сигналов, но и из-за присутствия шумов, помех, наводок из окружающей среды и двигательных артефактов, появляющихся в результате посторонней двигательной активности человека.

Чаще всего для усиления и фильтрации таких сигналов на практике используется архитектура аналогового входного каскада, изображенная на рис.1 [6].

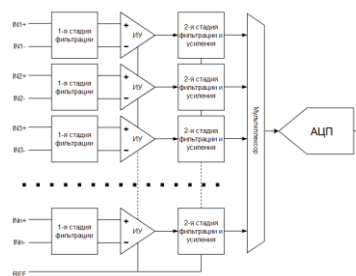


Рис.1. Стандартная архитектура многоканального усилителя биопотенциалов

Такая архитектура позволяет одновременно обрабатывать сигналы от множества электродов и имеет приемлемое соотношение отношение сигнал/шум (SNR). Но использование большого количества компонентов в стандартной архитектуре влечёт за собой значительное увеличение общего энергопотребления, что неприемлемо при разработке носимого устройства, где важным фактором является энергоэффективность. Кроме того, компактное размещение всех необходимых компонентов в устройстве может представлять некоторые технические трудности или вовсе быть невозможным, если только эта архитектура не воплощена в виде отдельного чипа.

Архитектура с быстрым мультиплексированием электродов для регистрации биопотенциалов

Для устранения этих проблем в целях минимизации размеров схемы предложена архитектура, изображенная на рис.2.

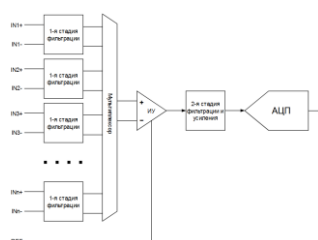


Рис.2. Предложенная архитектура многоканального усилителя биопотенциалов с использованием быстрого мультиплексирования электродов

Для сокращения общего количества компонентов первая стадия усиления и вторая стадия усиления и фильтрации объединены для всех входных сигналов. При этом также значительно снижается общее энергопотребление схемы.

Прежде такое решение использовалось для регистрации электрической активности мозга (электроэнцефалографии) в виде единой микросхемы [7].

В представленной архитектуре ключевую роль играет быстрое мультиплексирование электродов. Оно предполагает последовательное подключение входных электродов к одному усилительному каскаду, позволяя усилителю обрабатывать сигналы поочередно, один за другим.

В каждый промежуток времени через мультиплексор к усилителю подключается только один из входных каналов. Переключение между каналами происходит с высокой скоростью, что позволяет разным сигналам обрабатываться как бы одновременно.

Предложенный способ оцифровки мультиплексированного сигнала изображен на рис.3:

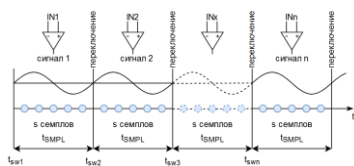


Рис.3. Вариант семплирования мультиплексированного сигнала микроконтроллером

Схемотехническое решение для усилителя биопотенциалов

В архитектуре с быстрым мультиплексированием электродов существуют две основные проблемы: переходный процесс мультиплексора и наличие паразитной емкости в переходе электрод-кожа, которая разряжается в момент после переключения.

Для решения этих проблем первая ступень фильтрации сигнала была расположена перед мультиплексором, чтобы защитить схему от вредных воздействий, уменьшить перекрестные помехи между каналами и нейтрализовать эффект переходного процесса.

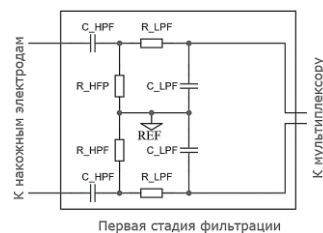
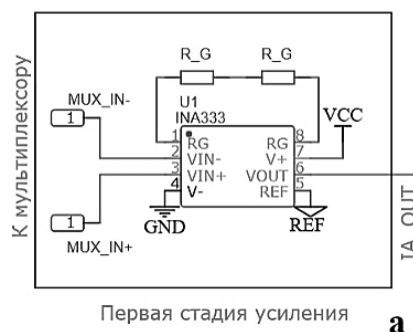


Рис.4. Первая стадия фильтрации сигнала

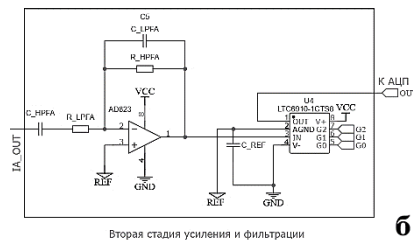
Первая стадия фильтрации сигнала в предложенной схеме состоит из двух RC-цепочек, образующих полосу частот (рис.4, $f_{HPF} = 1$ Гц и $f_{LPF} = 589$ Гц).

В качестве инструментального усилителя выбран INA333 с установленным коэффициентом усиления $G = 10$ (рис.5(a)).

Вторая стадия фильтрации включает активный полосовой фильтр на операционном усилителе (ОУ) AD823 с коэффициентом усиления 10 ($f_{HPF} = 40$ Гц и $f_{LPF} = 401$ Гц) и программно-управляемый усилитель (PGA) LTC6910-1 (рис.5(б)).



Первая стадия усиления **а**



Вторая стадия усиления и фильтрации **б**

Рис.5.(а) первая стадия усиления, (б) вторая стадия усиления на основе ОУ

Сборка схемы была произведена на нескольких печатных платах с использованием дополнительной безопасной макетной платы.

Полученные результаты

Для проверки работы архитектуры и схемы была произведена регистрация электромиограмм в двух режимах: одноканальный и двухканальный с быстрым мультиплексированием электродов. В первом режиме производилась запись электромиограммы лучевого сгибателя запястья, во втором – одновременная запись электромиограммы лучевого сгибателя запястья и плечелучевой мышцы. Использовались гелевые хлорсеребряные электроды ($d=10$ мм), расстояние между электродами $D=15$ мм. Полученные электромиограммы представлены соответственно на рис.6(а) и (б).

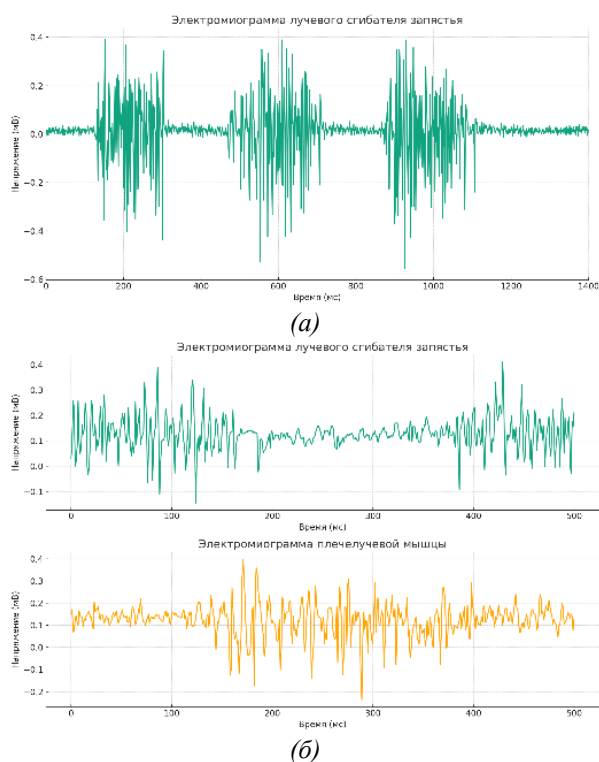


Рис.6. Полученные электромиограммы, (а) в одноканальном режиме, (б) в режиме быстрого мультиплексирования биопотенциалов

Для оценки качества работы усилителя биопотенциалов был рассчитано соотношение сигнал/шум (SNR). Значение RMS_{noise} для усилителя биопотенциалов равно 19 мкВ, а соотношение сигнал-шум $SNR = 18,7$ дБ, что является стандартным показателем для обычных систем регистрации поверхностной электромиографии и HD-sEMG [8].

Заключение

В этой работе исследовалось применение быстрого мультиплексирования электродов для создания компактной системы регистрации поверхностной электромиограммы высокой плотности.

Предложена архитектура быстрого мультиплексирования электродов. Для доказательства работы концепта был разработан и протестирован прототип устройства.

В дальнейшем планируется увеличить соотношение сигнал/шум (SNR), произведя сборку на единой печатной плате, добавить возможность быстрого восстановления (Fast Recovery, FR), а также использовать полученный усилитель биопотенциалов в носимом многоканальном биоинтерфейсе.

Список литературы

1. Kalc M. et al. Motor Unit Identification in the M Waves Recorded by High-Density Electromyogram // IEEE Trans Biomed Eng. IEEE Computer Society, 2023. Vol. 70, № 5. P. 1662–1672.
2. Hu X. et al. Finger Movement Recognition via High-Density Electromyography of Intrinsic and Extrinsic Hand Muscles // Sci Data. Nature Research, 2022. Vol. 9, № 1.
3. Proaño-Guevara D., Serpa-Andrade L. High Accuracy Multi-channel Surface EMG Acquisition System for Prosthetic Devices Control // Advances in Intelligent Systems and Computing. Springer, 2020. Vol. 1205 AISC. P. 125–130.
4. Electrophysiology products - EEG and EMG products - TMSi [Electronic resource]. URL: <https://www.tmsi.com/products/> (accessed: 01.02.2024).
5. Gary Kamen, David A. Gabriel. Essentials of Electromyography. Champaign, IL : Human Kinetics, 2010. 133–137 p.
6. Koutsoftidis S. et al. Myolink: A 128-Channel, 18 nV/ $\sqrt{\text{Hz}}$, Embedded Recording System, Optimized for High-Density Surface Electromyogram Acquisition // IEEE Trans Biomed Eng. IEEE Computer Society, 2022. Vol. 69, № 11. P. 3389–3396.
7. Sharma M., Strathman H.J., Walker R.M. Verification of a Rapidly Multiplexed Circuit for Scalable Action Potential Recording // IEEE Trans Biomed Circuits Syst. NLM (Medline), 2019. Vol. 13, № 6. P. 1655–1663.
8. Fraser G.D. et al. Automated Biosignal Quality Analysis for Electromyography Using a One-Class Support Vector Machine // IEEE Trans Instrum Meas. 2014. Vol. 63, № 12. P. 2919–2930.

Секция «Электронное приборостроение»

СИСТЕМА АВТОМАТИЧЕСКОЙ ЗАМЕНЫ АККУМУЛЯТОРА БПЛА

Козаев З.З., Посмитный С.В., Старилова Е.В.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Данная работа посвящена разработке программно-аппаратного комплекса, решающего проблему ограниченного времени работы беспилотных летательных аппаратов (БПЛА). Предлагаемый комплекс включает в себя систему автоматической смены и зарядки аккумуляторов, обеспечивая тем самым возможность более длительных полетов и эффективного использования дронов в различных областях.

Введение

С развитием технологий и расширением применения беспилотных летательных аппаратов (БПЛА) в современном обществе возникает задача увеличения их автономности. В данной работе представлен программно-аппаратный комплекс, целью которого является увеличение длительности полета дронов с помощью механизма автоматической смены и зарядки аккумуляторных батарей. Проект ставит перед собой несколько ключевых задач: разработку механизма смены батареи, проектирование прототипа станции и модуля для дрона, используя 3D-печать, сервоприводы [1] и микроконтроллеры.

Предполагаемые области применения данной технологии:

- Увеличение длительности работы и дальности полета при доставке дронами грузов.
- Осуществление бесперебойной работы в труднодоступных для человека местах.
- Поддержание автономности дрона в системах наблюдения и контроля.

Существуют несколько корейских патентов на станцию для быстрой замены аккумулятора дрона: KR101705838B1 и KR20190055925A. Каждый из них имеет существенные недостатки, которые будут доработаны в данной работе. В первом патенте [2] – дрон осуществляет приземление в 4 углубления для 4 ножек. В реальной же ситуации это невозможно, так как при посадке дрон имеет смещение в позиции и посадить его четко в углубления нельзя. Далее, аккумулятор не прикреплен к корпусу дрона, а размещен на плате контактами вниз. При любом колебании в полете есть вероятность его потери. Второй патент [3] уже имеет крепление аккумулятора, однако не включает в себя процесс позиционирования дрона после посадки и контакты для перезарядки батареи. В нем описан только процесс смены аккумуля

тора, что затрудняет его выполнение в реальной жизни.

Цель работы: Создание станции быстрой замены аккумулятора для увеличения автономности дрона (БПЛА) и дальности полета путем автоматической смены аккумуляторной батареи.

Описание станции и принцип работы

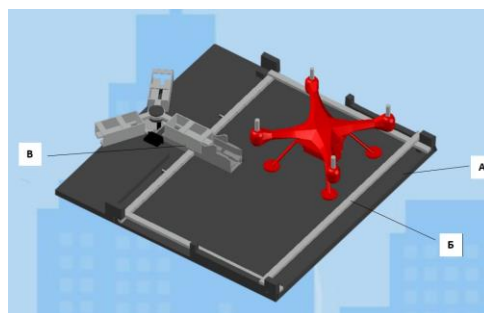


Рис.1. 3D модель станции

Станция для быстрой замены аккумулятора дрона состоит из:

- а) посадочной площадки
- б) механизма позиционирования
- в) системы замены

Помимо станции, необходим ещё специальный модуль к дрону. Данный модуль включает крепеж (рис.2А), содержащий дополнительный аккумулятор и электронные компоненты, коробку (рис.2Б) с крышкой (рис.2В), в которой располагается основной аккумулятор. Дополнительный аккумулятор нужен для поддержания бесперебойной работы дрона. На крышке коробки располагаются два контакта из нержавеющей стали [4]

Все необходимые компоненты распечатаны на 3D-принтере и доработаны с помощью пайки.

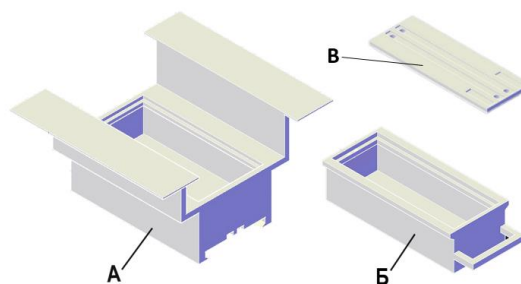


Рис.2. Прототип сменного модуля

Механизм позиционирования дрона

Механизм позиционирования в своём составе имеет: две продольные планки, одну поперечную планку, сервоприводы постоянного вращения, шаговые ремни, плату Arduino, соединительные провода

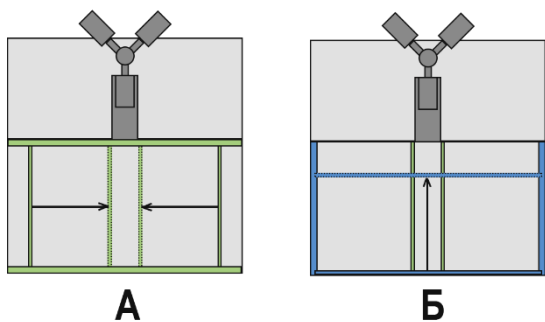


Рис.3. Принцип работы механизма позиционирования дрона

После приземления дрона на посадочную площадку (отдельное место на самой станции), сервоприводы постоянного вращения, после подачи сигнала с Arduino, приводят в действие шаговые ремни, на которых закреплены планки сдвига. Схема процесса представлена на рис.3. Планки сдвига взаимодействуют с дроном, перемещая его в нужную точку для дальнейшего процесса смены аккумуляторной батареи.

Две продольные планки смещают дрон к середине станции (рис.3А), поперечная – пододвигает дрон к устройству смены аккумулятора (рис.3Б).

Устройство замены батарей

Система замены состоит из двух компонентов – основной части для замены аккумулятора и «мельницы», осуществляющей зарядку и выдачу батарей.

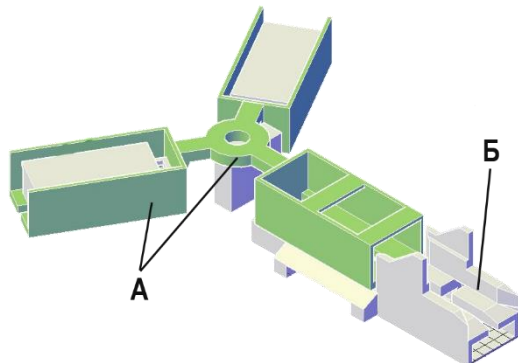


Рис.4. Прототип устройства замены

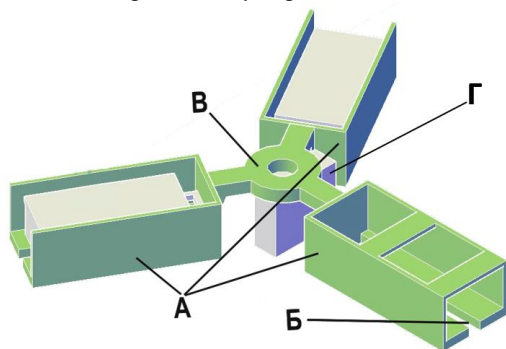


Рис.5. Механизм «Мельница»

«Мельница» состоит из трех контейнеров (рис.5А) с проемами (рис.5Б) В каждом из них предусмотрена пара контактов для зарядки аккумуля-

лятора. Все три контейнера сцеплены деталью (рис.5В), которая крепится на сервопривод (рис.5Г), управляемый Arduino. Важно отметить, сервопривод имеет максимальный угол поворота 270 градусов, что позволяет задать три позиции – 0 градусов, 120, 240. Сервоприводов, способных осуществлять поворот на 360 градусов с возможностью задавать определенный угол не существует. Сервопривод постоянного вращения не подходит по той же причине, так как имеет люфт при повороте.

Прототип устройства для фиксации дрона и вытягивания аккумуляторной батареи (рис.4Б) представляет собой деталь, состоящую из двух частей и механизма для сдвига батареи. Первая часть (рис.7, часть 1) – небольшая площадка со скосами и выступами для фиксации дрона на ней. Скосы (рис.7А) позволяют разместить дрон на платформе без заклиниваний даже при наличии небольшого сдвига при его позиционировании. Вертикальные выступы (рис.7Б) необходимы для блокирования корпуса дрона при вытягивании аккумулятора.

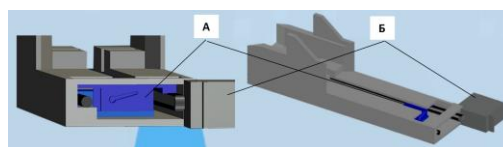


Рис.6. Устройство вытягивания аккумулятора

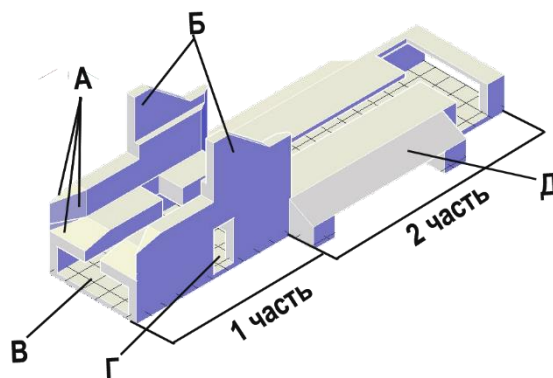


Рис.7. Основная деталь устройства вытягивания аккумулятора

На устройстве вытягивания аккумулятора имеется прорезь для хода в нем механизма (рис.7В), состоящего из сервопривода постоянного вращения, шагового ремня, металлической планки и сервопривода с позиционированием на угол 90 градусов. Отверстие на рис.7Г позволяет сервоприводу поворачиваться, не упираясь в борта корпуса при его движении.

Принцип работы заключается в следующем: дрон с помощью механизма позиционирования размещается в 1 части основной детали, фиксируя свой корпус. Сервопривод (рис.6А), находящийся на уровне с отверстием Г осуществляет поворот на 90 градусов, тем самым цепляя корпус аккумулятора за специальную ручку. Сервопривод постоянного вращения (рис.6Б) сдвигает ремень и с ним аккумулятор и сервопривод (рис.6А) во вторую часть детали. В ней также предусмотрено отверстие по всей длине для

передвижения сервопривода (рис.6А) без препятствий. Скосы по бокам (рис.7Д) предназначены для беспрепятственного вращения мельницы во второй части детали. Сервопривод (рис.6А) стаскивает в один из ее слотов аккумулятор и отцепляет его.

В работе станции и дрона существуют автоматический и ручной режимы. В ручном режиме производится посадка БПЛА и активация станции. Следующий этап – запуск процесса позиционирования дрона и замены аккумуляторной батареи. Данный процесс происходит автоматически.

Заключение

Таким образом, разработан прототип станции, производящий позиционирование и замену аккумуляторной батареи у дрона. В ходе данной разработки были исправлены недостатки аналогичных устройств – добавлен механизм позиционирования, создан универсальный крепеж для аккумулятора и удобный корпус, спроектирован механизм «мельница», заряжающий и выдающий готовые к работе батареи.



Рис.8. Станция быстрой замены аккумулятора дрона

На рис.8 показана станция на ранних этапах реализации проекта. Система позиционирования на данный момент разрабатывается.

В конечном итоге, мы получили станцию, которая будет иметь возможность перезарядить дрон при низком уровне заряда его батареи. Это позволит значительно увеличить дальность полета дрона при наличии нескольких таких устройств на пути его следования.

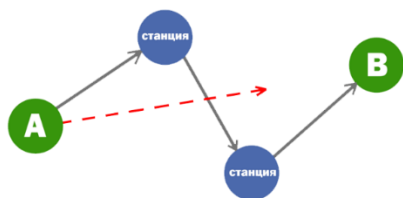


Рис.9. Увеличение дальности полета дрона

Где А и В на рис.9 начало и конец пути соответственно, а станции – промежуточные пункты зарядки аккумулятора с целью увеличения дальности полета и автономности дрона.

В будущем планируется продолжение работы над данной темой. В частности, создание возможности осуществления автоматической посадки и начала

работы станции замены; обеспечение защиты от случайного или намеренного повреждения, кражи батарей, погодных условий; разработка системы распознавания свой–чужой, без чего практическая реализация невозможна.

Список литературы

1. В. Ф. Козаченко, А. А. Жарков, Д. И. Савкин «Цифровой сервопривод с шаговыми и вентильными двигателями: учебник для вузов», Москва НИУ МЭИ 2022
2. Патент «Automatic battery exchange system for drone»
[https://patents.google.com/patent/KR101705838B1/en?q=\(drone+battery+automatic+replacement+station\)&oq=drone+battery+automatic+replacement+station](https://patents.google.com/patent/KR101705838B1/en?q=(drone+battery+automatic+replacement+station)&oq=drone+battery+automatic+replacement+station)
3. Патент «Battery exchange method and system for drone»
[https://patents.google.com/patent/KR20190055925A/en?q=\(drone+battery+replacement+station\)](https://patents.google.com/patent/KR20190055925A/en?q=(drone+battery+replacement+station))
4. Бредихин А.Н. Хомяков, М.В. «Электрические контактные соединения», Москва, 1980.
5. Джереми Блум, «Изучаем Arduino и инструменты и методы технического волшебства», Санкт-Петербург «БХВ-Петербург», 2017.
6. Учебно–методическое пособие. «Компьютерное конструирование в AutoCAD 2016. Начальный курс». Ф.А. Перепелица. Санкт–Петербург 2015.

ПРИЁМ, ПЕРЕДАЧА И РАСШИФРОВКА СООБЩЕНИЙ С ПЕРЕДАТЧИКОВ ADS-B С ПОМОЩЬЮ МАЛЫХ КОСМИЧЕСКИХ АППАРАТОВ

Елькин В.А., Мотайленко И.А.,
Чернявских И.И., Носов И.А.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В данной статье исследуется возможность использования малых космических аппаратов для приема, передачи и декодирования сообщений с транспондеров ADS-B. Описывается процесс работы ADS-B, причины, обусловившие потребность в разработке подобных технологий, и, кроме того, преимущества и ограничения использования малых космических аппаратов для данных целей, в частности, реализация данного функционала при взаимодействии с малым космическим аппаратом CubeSХ-HSE 3.

Введение

ADS-B (Automatic Dependent Surveillance-Broadcast) – это метод передачи информации о полете самолета, который использует встроенные в воздушные суда автономные передатчики [1]. Его целью является повышение эффективности и безопасности

воздушного движения. Основной принцип работы этого метода заключается в том, что самолет передает информацию о своем местоположении, скорости, высоте и других параметрах через специализированную радиосхему. Эти данные затем передаются по радиоканалу транспондера на наземные станции управления воздушным движением и другим самолетам, находящимся в радиусе действия.

Следует отметить, что у ADS-B есть определенные ограничения. В частности, для полноценной работы системы необходимо, чтобы все самолеты, осуществляющие полеты, были оборудованы данными транспондерами, иначе возможны проблемы с обнаружением и предупреждением о приближающихся воздушных судах. Кроме того, у ADS-B имеет ограниченную дальность передачи данных.

Обзор существующих решений

В настоящее время несколько компаний, в частности, Aelen Space и Aireon, занимаются установкой системы ADS-B на космических аппаратах. Значительные успехи продемонстрировала компания Aireon, развернувшая сеть спутников с транспондерами ADS-B на базе спутников Iridium NEXT. Эта сеть обеспечивает глобальное покрытие для контроля воздушного движения с использованием данной технологии. Как правило, представленные на рынке решения не основываются на использовании малых космических аппаратах для создания более универсальных решений.

Малый космический аппарат (МКА) CubeSX-HSE 3, запущенный в качестве попутной нагрузки на ракете-носителе "Союз", оборудован транспондером ADS-B. С его помощью студенты могут отслеживать текущее местоположение воздушных судов и изучать методы передачи и шифрования данных. В Московском институте электроники и математики им. А. Н. Тихонова Национального исследовательского университета «Высшая школа экономики» в настоящее время студенты разрабатывают специальное программное обеспечение, предназначенное для отображения в реальном времени данных с транспондера ADS-B, полученных со спутника [2-4].

В ходе взаимодействия с данным спутником было выявлено, что значительная часть сообщений ADS-B сильно повреждается помехами, что приводит к потере большого количества данных. Именно эта проблема стала основным стимулом для поиска альтернативного способа обработки получаемых сообщений.

Структура сообщения

Сообщение ADS-B имеет длину 112 бит и состоит из пяти основных частей, показанных следующим образом:

Таблица 1. Структура сообщения

| | | | | |
|--------|--------|-----------|---------|---------|
| DF (5) | CA (3) | ICAO (24) | ME (56) | PI (24) |
|--------|--------|-----------|---------|---------|

Сообщение ADS-B для гражданских летательных аппаратов начинается с формата нисходящей линии связи (DF) 17. Биты 6–8 указывают возможности транспондера (CA), после находится 24-битный код

транспондера (известный как код ICAO). Последние два сегмента представляют собой 56-битную полезную нагрузку (ME) и 24-битную четность (PI).

Для дешифровки ME-части сообщения нам нужно взглянуть на код типа сообщения, расположенный в первых 5 битах сегмента ME. В таблице 2 показаны связи между каждым кодом типа и его информацией, содержащейся в сегменте ME.

Таблица 2. Соответствие кодов типа и содержимого сообщений

| Код типа | Содержимое данных |
|----------|---|
| 1–4 | Идентификация летательного аппарата |
| 5–8 | Положение относительно поверхности |
| 9–18 | Положение в воздухе (с высотой в Baro) |
| 19 | Скорость в воздухе |
| 20–22 | Положение в воздухе (с высотой в GNSS) |
| 23–27 | Зарезервированные |
| 28 | Статус летательного аппарата |
| 29 | Состояние и информация о статусе |
| 31 | Статус эксплуатации летательного аппарата |

Обзор метода расшифровки данных

Отметим, что данный протокол является открытым, и, как следствие, при передаче данных не требуется какое-либо дополнительное шифрование помимо стандартного. Сообщение о местоположении и высоте самолета содержит код типа, принимающий значения в диапазонах [9;18] или [20;22]. Закодированные значения широты и долготы не являются фактическими значениями, а передаются в формате Compact Position Reporting (CPR), который использует меньше битов для более точного кодирования положения. Сообщения передаются последовательно и могут быть дешифрованы двумя разными способами определения местонахождения в воздухе.

1) **Глобальное однозначное декодирование позиции:** не зная опорную точку декодирует местоположение летательного аппарата.

2) **Локально однозначное декодирование позиции:** зная опорную позицию использует только одного сообщения для декодирования.

Для вычисления местонахождения МКА используется технологии TLE и SGP4. TLE (Two-Line Element Set) - формат описания орбитальных параметров и положения искусственного спутника Земли. Первая строка содержит название спутника и номер строки, вторая - орбитальные параметры. Используя полученные данные и алгоритм расчета SGP4 (аналитическое решение задачи движения искусственного спутника по эллиптической орбите), можно вычислить положение МКА в выбранное время. Расчеты можно проводить в разных системах координат, обычно используются геодезические/ геоцентрические координаты. TLE-данные являются математической моделью орбиты спутника и могут иметь погрешность.

Благодаря локально-однозначному декодированию возможно использовать только одного сообщения для декодирования. Данный метод позволяет вычислить индекс широты (j) и индекс долготы (m) на основе местоположения спутника в момент приема сигнала и может применяться как к нечетным, так и к четным сообщениям. Существует ограничение диапазона в 180 морских миль, предназначенное для обеспечения согласованности зон широты и долготы между опорной позицией и позицией декодирования. В свою очередь, из-за размеров МКА, а также конструктивных особенностей антенны, радиус приема данных не превосходит 180 морских миль. Следовательно, в случае получения сообщения, МКА всегда находится в пределах 180 морских миль относительно самолёта, что позволяет использовать локально однозначное декодирование. Исходное положение МКА обозначено как lat_r , lon_r , а положение судна - lat_c , lon_c .

Для вычислений понадобятся несколько основных функция и понятий. N_z – количество зон широты между экватором и полюсом, в данном случае 15. $floor(x)$ возвращает наибольшее целочисленное значение k , где $k \leq x$. Функция $mod(x, y)$ определена формулой (1). $NL(lat)$ номер зоны долготы (2).

$$mod(x, y) = x - y \times floor\left(\frac{x}{y}\right), y \neq 0 \quad (1)$$

$$NL(lat) = floor\left\{\frac{2\pi}{arccos\left[1 - \frac{1 - \cos\left(\frac{\pi}{2N_z}\right)}{\cos^2\left(\frac{\pi}{180} \times lat\right)}\right]}\right\} \quad (2)$$

Для расчета широты (5) требуется присвоить переменной i значение бита F , так как размер зоны широты (3) различается в зависимости от типа сообщения, далее требуется найти индекс широты (4).

$$dLat = \frac{360}{4n_z - i} \quad (3)$$

$$j = floor\left(\frac{lat_r}{dLat}\right) + floor\left[\frac{mod(lat_r, dLat)}{dLat} - lat_c + \frac{1}{2}\right] \quad (4)$$

$$lat = dLat \times (j + lat_c) \quad (5)$$

Следующим шагом является вычисление приращение долготы (6) для каждой зоны на основе декодированной широты, которая зависит как от типа сообщения, так и от широты. Зная приращение можно вычислить индекс долготы (7) и в следствии чего долготу (8).

$$dLon = \frac{360}{max(NL(lat) - i, 1)} \quad (6)$$

$$m = floor\left(\frac{lon_r}{dLon}\right) + floor\left(\frac{mod(lon_r, dLon)}{dLon} - lon_c + \frac{1}{2}\right) \quad (7)$$

$$lon = dLon \times (m + lon_c) \quad (8)$$

Заключение

Использование местоположения МКА для однозначного декодирования сообщений о положении летательного аппарата позволяет расшифровать его, используя только одно сообщение. Данный метод является эффективным с точки зрения затрачиваемо-

го времени и, очевидно, объёма данных, требуемых для нахождения геолокации воздушного судна, что расширяет потенциальный спектр возможностей для его применения. Кроме того, использование данного подхода позволяет уменьшить риск потери и/или искажения части передаваемых сообщений, что также положительно сказывается на последующей обработке полученных данных.

Список литературы

1. Junzi S. The 1090 megahertz riddle, second edition / S. Junzi - TU Delft OPEN Publishing, 2021. - 160 p.

2. Страница проекта по расшифровке сообщений с ADS-B в электронном каталоге проектов МИЭМ [Электронный ресурс]. – Режим доступа: <https://cabinet.miem.hse.ru/#/project/1689/> (25.10.2023)

3. Baleskin V.A. Virtual space virtual satellite / V.A. Baleskin, M.A. Bubnova, A.N. Kosinov, D.A. Abrameshin, L.A. Romanov, B.S. Nikitin // Journal of Physics: Conference Series. - 2021. - vol. 2134. [Electronic resource]. - Access mode: <https://publications.hse.ru/articles/545013185> (24.09.2023)

4. Abrameshin D., First Earth-Imaging CubeSat with Harmonic Diffractive Lens / D. Abrameshin, N. Ivliev, V. Evdokimova, V. Podlipnov, M. Petrov, S. Ganchevskaya, I. Tkachenko, Y. Yuzifovich, A. Nikonorov, R. Skidanov, N. Kazanskiy, V. Soifer // Electronics (MDPI). - 2022. - vol. 14.(9). [Electronic resource]. - Access mode: <https://publications.hse.ru/articles/623248383> (16.09.2023).

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ РАСЧЕТНОЙ ОЦЕНКИ ПОКАЗАТЕЛЕЙ БЕЗОТКАЗНОСТИ ИНФОРМАЦИОННОЙ СИСТЕМЫ НАЗЕМНОГО КОМПЛЕКСА УПРАВЛЕНИЯ РОЕМ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Костюк А.А., Старкова Э.А.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В данной работе был проведён анализ моделей и методов оценки надёжности программного кода и технического оборудования. На основе полученных данных построена и применена математическая модель оценки вероятности безотказной работы (ВБР) информационной системы наземного комплекса управления (ИС НКУ) роём беспилотных летательных аппаратов (БПЛА) в период нормальной работы.

Введение

В настоящее время ужесточаются требования к показателям надёжности, появляется необходимость в уточнении их оценки в такого рода системах.

Надёжность информационных (вычислительных) систем состоит из совокупности надёжности технического оборудования и программного обеспечения [1].

В рое БПЛА, как и в любой современной инфокоммуникационной системе, ИС играет важную роль. Так, основной сервер и компьютеры находятся на центральной станции управления, непосредственно связанной со шлюзовым БПЛА (БПЛА-Лидером) в классической архитектуре построения системы связи роа, а сам программный продукт является «мозгом» системы, управляет полётом и следит за выполнением задач полёта.

В настоящей работе предлагается модель оценки показателей безотказности ИС НКУ роом БПЛА. В первых двух разделах анализируются базовые методы расчёта надёжности восстанавливаемых систем и ПО. В третьем и четвёртом разделе создаётся и применяется математическая модель.

Анализ существующих методов оценки надёжности аппаратной части ИС НКУ

Основными техническими средствами (ТС) информационных систем НКУ являются (см. рис.1) [2]:

1. Управляющая ЭВМ (УЭВМ) (составные части: монитор, системный блок, мышь и клавиатура).
2. Портативная персональная ЭВМ (ППЭВМ) (составные части: монитор, системный блок и пульт ручного дистанционного управления).
3. Пульт индикации и выбора режимов (ПИВР).
4. Приёмо-передающая аппаратура командно-телеметрического канала (ППАКТК).
5. Сетевая инфраструктура (СИ) (составные части: модем, коммутаторы, маршрутизаторы и система кабелей).

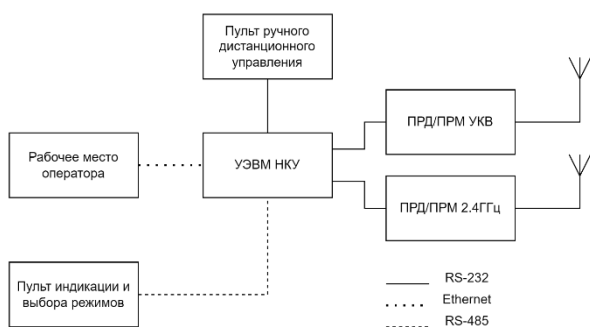


Рис.1. Структурная схема ИС НКУ

Информационные системы являются программно-аппаратным комплексом и относятся к восстанавливаемым системам. Рассмотрим и проанализируем базовые методы оценки надёжности восстанавливаемых систем [3].

Одним из основных методов оценки показателей безотказности технических средств является усечённый нормальный закон надёжности, отражающий вероятность отказа ТС при длительном и постепенном изменении его характеристик, то есть при его старении и износе. Поскольку усечённый нормаль-

ный закон надёжности используется для указанного выше периода, а в данном тезисе рассматривается период нормальной работы ТС НКУ, который соответствует постоянной интенсивности отказов, усечённый нормальный закон использовать нецелесообразно.

Следующий проанализированный закон – закон распределения Релея так же применим в случаях отказов системы в результате её износа или старения, вследствие чего его дальнейшее использование в работе нецелесообразно.

Рассмотрим закон Пуассона. Он является дискретным законом и используется, как правило, для определения вероятности возникновения k отказов за определённый промежуток времени t , что применимо для восстанавливаемых систем, однако в работе рассматривается наработка информационной системы до первого отказа, следовательно, закон Пуассона использовать нецелесообразно.

Экспоненциальный закон надёжности применяется, когда отказы ТС случайны и независимы, что в большинстве случаев выражается в период нормальной работы аппаратуры. Для данного закона ВБР имеет зависимость только от величины исследуемого интервала Δt , а прошедшее время не учитывается. При данном законе интенсивность отказов является постоянной, то есть $\lambda(t) = \lambda = const$. Ниже приведены некоторые формулы, описывающие экспоненциальный закон надёжности (1 – 3):

$$f(t) = \begin{cases} \lambda * \exp(-\lambda t), & t > 0; \\ 0, & t \leq 0; \end{cases} \quad (1)$$

$$P(t) = \exp(-\lambda t); \quad (2)$$

$$Q(t) = 1 - \exp(-\lambda t); \quad (3)$$

где $f(t)$ – частота отказов технической системы, $P(t)$ – вероятность безотказной работы, $Q(t)$ – вероятность отказа.

Проведённый анализ показал, что экспоненциальный закон надёжности является достаточно универсальным для периода нормальной работы технических средств информационной системы наземного комплекса управления роом БПЛА и часто используется для класса информационно-управляющих систем. Поэтому целесообразно при расчётах надёжности ИС НКУ использовать этот закон.

Анализ существующих моделей оценки надёжности программной части ИС НКУ

Существует множество моделей, способных оценить надёжность программного обеспечения, используя различные подходы. Самые популярные и точные модели будут рассмотрены в этом разделе. Для наглядности представлена классификация моделей надёжности программного обеспечения (см. рис.2).

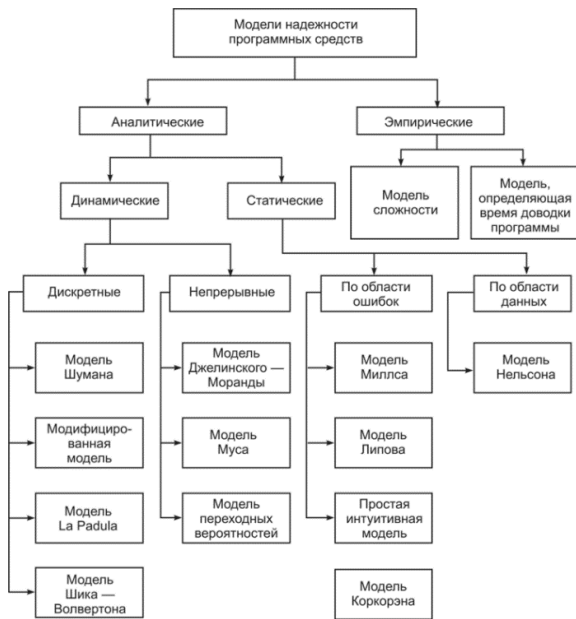


Рис.2. Классификация моделей надёжности ПО

Рассмотрим модель Муса, где решение по оценке показателей безотказности базируется на времени работы кода до возникновения его очередного отказа. В данной модели лежит предположение о том, что существуют ошибки, которые не приводят к отказу всей системы в целом, таким образом, на интервале наблюдения может появляться несколько ошибок. Для оценки показателей безотказности по модели Муса требуется получить исходные данные с помощью иных моделей, что приводит к дополнительным временным затратам, поэтому её использование нецелесообразно.

Проанализируем модель Шумана на целесообразность её использования в данной работе. Тестирование программного кода в данной модели подразумевает несколько шагов. На первом шаге проводится серия тестовых испытаний, в которой выявляются ошибки. Их исправление предполагается только по завершении специально отведённых тестов. При необходимости комплекс тестов корректируется перед началом следующего этапа. Таким образом, предполагается, что количество ошибок в исходном коде фиксировано и уменьшается по мере их выявления и исправления, то есть новые ошибки не вносятся. Модель является достаточно универсальной, однако данная модель не учитывает, что при корректировке ПО могут быть внесены новые ошибки, что снижает возможность её использования для оценки надёжности ПО НКУ. В настоящее время существуют более рациональные модели, к которым относится модель Джелински-Моранды и её улучшенная модификация Шика-Волвертона.

Перейдём к модели Джелински-Моранды. В данной модели измеряются интервалы времени между возникновением двух ошибок в программе. Полученные значения имеют экспоненциальное распределение с интенсивностью отказов, которая пропорциональна количеству оставшихся (не выявленных)

ошибок. Найденная ошибка исправляется, а число оставшихся, соответственно, уменьшается.

Частота отказов и ВБР ПО в модели Джелински-Моранды вычисляются по следующим формулам:

$$f(t_i) = \lambda_i * \exp(-\lambda_i t_i), \quad (4)$$

$$P(t_i) = \exp(-\lambda_i t_i) \quad , \quad (5)$$

где интенсивность отказов:

$$\lambda_i = C * (N - i + 1), \quad (6)$$

где C – коэффициент пропорциональности, N – начальное количество ошибок программы.

Параметры C, N находят посредством измерения времени до очередного отказа t_1, t_2, \dots, t_k при тестировании программы и последующего решения системы уравнений:

$$\begin{cases} \sum_{i=1}^k \frac{1}{N - i + 1} = \frac{K}{N + 1 - Q * K}, \\ C = \frac{K}{A(N + 1 - Q * K)}, \end{cases} \quad (7)$$

где

$$Q = \frac{B}{A * K}, \quad (8)$$

$$A = \sum_{i=1}^k t_i, \quad (9)$$

$$B = \sum_{i=1}^k i * t_i, \quad (10)$$

Данная модель не учитывает, что при неправильном определении начального числа ошибок в программном коде интенсивность отказов ПО может получиться отрицательной по формуле (6), что приведёт к бессмысленному результату. При этом данная модель отвечает нужным требованиям по поиску вероятности безотказной работы ПО НКУ в течение большого промежутка времени, то есть её использование в данной тематике целесообразно.

Модель Шика-Волвертона является модификацией модели Джелински-Моранды. По данной модели существует вероятность возникновения более одной ошибки на исследуемом интервале. Частота возникновения ошибок пропорциональна как количеству ошибок в коде, так и времени теста, таким образом, вероятность нахождения ошибки увеличивается с течением времени. Данная модель даёт наилучшие результаты при её использовании для крупномасштабных проектов. Однако по данной модели сложно оценить ВБР ПО на долгом промежутке времени, поскольку она отвечает на вопросы о вероятности возникновения каждой последующей ошибки.

Разработка единой математической модели

Рассмотрим информационную структуру НКУ для оценки её надёжности.

Поскольку при отказе каждой из частей системы связь с роум БПЛА будет потеряна, что приведёт к невыполнению поставленной задачи, структурная схема надёжности имеет последовательное соединение частей ИС наземного комплекса управления (см. рис.3).

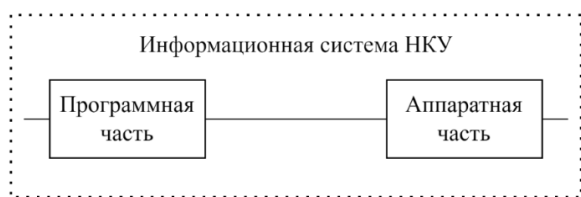


Рис.3. Структурная схема надёжности ИС наземного комплекса управления роум БПЛА

При последовательном соединении элементов системы ВБР вычисляется по следующей формуле:

$$P(t) = \prod_{i=1}^n P_i(t). \quad (11)$$

Таким образом, с учётом предложенных моделей, вероятность безотказной работы ИС НКУ будет вычисляться по формуле, которая представляет собой композицию:

$$P(t) = \exp(-\lambda_1 t) * \exp(-\lambda_2 t), \quad (12)$$

где λ_1 – интенсивность отказов ТС НКУ, λ_2 – интенсивность отказов ПО, определяемая по модели Джелински-Моранды по формуле (6).

Применение разработанной математической модели

Для примера оценки надёжности программного обеспечения ИС НКУ по модели Джелински-Моранды взято $n = 65$ ошибок за 2 года (730 дней) на этапе отладки. Дни X_i между появлением ошибок i распределены случайным образом.

Для расчёта надёжности такого ПО была составлена таблица (см. рис.4).

| i | X_i | $i * X_i$ | m | $g_n(m, A)$ | $f_n(m)$ | $f_n(m) - g_n(m, A)$ |
|----|----------|-----------|-----|-------------|----------|----------------------|
| 1 | 18 | 18 | 66 | 2,002194 | 4,759276 | 2,757081595 |
| 2 | 2 | 4 | 67 | 1,942364 | 3,774427 | 1,832063209 |
| 3 | 9 | 27 | 68 | 1,886005 | 3,289352 | 1,403346797 |
| | | | ... | | | |
| 63 | 22 | 1386 | 128 | 0,68809 | 0,712942 | 0,02485195 |
| 64 | 17 | 1088 | 129 | 0,680882 | 0,704881 | 0,023998769 |
| 65 | 10 | 650 | 130 | 0,673824 | 0,697008 | 0,023184149 |
| | 730 | 24481 | | | | |
| A | | | | | | |
| | 33,53562 | | | | | |

Рис.4. Таблица для расчёта надёжности ПО по модели Джелински-Моранды

где m – число прогнозируемых ошибок

$$m \geq n + 1, \quad (13)$$

$$g_n(m, A) = \frac{n}{m - A}, \quad (14)$$

$$A = \frac{\sum_{i=1}^n i * X_i}{\sum_{i=1}^n X_i}, A \geq \frac{n + 1}{2}, \quad (15)$$

$$f_n(m) = \sum_{i=1}^n \frac{1}{m - i}, \quad (16)$$

Представленные выше выражения (14, 16) получаются путём упрощения формулы для нахождения начального количества ошибок из системы (7). По полученным данным в таблице анализируется разность функций $f_n(m) - g_n(m, A)$ и производится оценка максимального правдоподобия для начального числа ошибок N (выбирается наиболее близкое к нулю значение разности функций).

Таким образом, в представленной системе $N = m - 1 = 130 - 1 = 129$. Далее по формуле

$$C = \frac{n}{(N + 1) * \sum_{i=1}^n X_i - \sum_{i=1}^n i * X_i}, \quad (17)$$

вычисляется коэффициент пропорциональности, в данном примере он равен 0,000923.

Итак, интенсивность отказов λ в представленном примере оказывается равной 0,059075 по формуле (6). Соответственно, вероятность безотказной работы данного ПО через 2 года окажется равной 0,942636 по формуле (5).

Средняя наработка на отказ системного блока (СБ) в настоящее время обычно составляет 45000 ч, монитора (МН) DURApixel – 100000 ч, клавиатуры (К) – 100 млн срабатываний (833333 ч), мыши (М) 1600DPI – 10 млн кликов (1666667 ч), пульта ручного дистанционного управления (ПРДУ) – 40000 ч, кабельной системы (КС) ТУ 16.К71-214-94 – 100000 ч, модема (МД) ONCELL G3470A-LTE-EU – 327326 ч, коммутатора (КМ) Cisco C9300L-24T-4G-A – 395800 ч, маршрутизатора (МР) Cisco ISR4221-SEC/K9 – 593270 ч, приёмно-передающей аппаратуры командно-телеметрического канала – 10000 ч, пульта индикации и выбора режимов – 40000 ч.

Согласно структурной схеме ИС НКУ (рис.1) в системе отсутствует резервирование, таким образом система соединена последовательно, вероятность безотказной работы аппаратной части информационной структуры будет вычисляться по экспоненциальному закону, выбранному ранее, следующим образом:

$$\exp(-\lambda_1 t) = \exp(-\lambda_{уэвм} - \lambda_{ппэвм} - \lambda_{пивр} - \lambda_{ппактк} - \lambda_{си}) * t, \quad (18)$$

Также, отказ любой составной части приведёт к отказу оборудования, поэтому интенсивность отказов аппаратуры равна сумме интенсивностей отказов её составных частей:

$$\lambda_{уэвм} = \lambda_{СБ} + \lambda_{МН} + \lambda_{К} + \lambda_{М}, \quad (19)$$

$$\lambda_{ппэвм} = \lambda_{СБ} + \lambda_{МН} + \lambda_{ПРДУ}, \quad (20)$$

$$\lambda_{СИ} = \lambda_{КС} + \lambda_{МД} + \lambda_{КМ} + \lambda_{МР}, \quad (21)$$

Интенсивность отказов в период нормальной эксплуатации является постоянной величиной и связана со средней наработкой на отказ каждого элемента ПК \bar{T} следующим соотношением:

$$\bar{T} = \frac{1}{\lambda}, \quad (22)$$

Для удобства при расчёте интенсивности отказов аппаратной части ИС НКУ по формулам (19-22) в таблице 1 представлены вышеперечисленные данные по средней наработке на отказ.

Таблица 1. Средняя наработка на отказ и соответствующая интенсивность отказов технического оборудования

| Средняя наработка на отказ аппаратуры ИС НКУ, ч | | | Интенсивность отказов, $\frac{1}{ч}$ |
|---|--|---------|--------------------------------------|
| УЭ ВМ | Средняя наработка на отказ составных частей, ч | | $\lambda_{УЭВМ} = 3,4 * 10^{-5}$ |
| | СБ | 45000 | $\lambda_{СБ} = 2,22 * 10^{-5}$ |
| | МН | 100000 | $\lambda_{МН} = 1 * 10^{-5}$ |
| | К | 833333 | $\lambda_{К} = 1,2 * 10^{-6}$ |
| | М | 1666667 | $\lambda_{М} = 6 * 10^{-7}$ |
| ПП ЭВ М | СБ | 45000 | $\lambda_{СБ} = 2,22 * 10^{-5}$ |
| | МН | 100000 | $\lambda_{МН} = 1 * 10^{-5}$ |
| | ПР ДУ | 40000 | $\lambda_{ПРДУ} = 2,5 * 10^{-5}$ |
| СИ | КС | 100000 | $\lambda_{КС} = 1 * 10^{-5}$ |
| | МД | 327326 | $\lambda_{МД} = 3,06 * 10^{-6}$ |
| | КМ | 395800 | $\lambda_{КМ} = 2,53 * 10^{-6}$ |
| | МР | 593270 | $\lambda_{МР} = 1,69 * 10^{-6}$ |
| ППАКТК | | 10000 | $\lambda_{ППАКТК} = 1 * 10^{-4}$ |
| ПИВР | | 40000 | $\lambda_{ПИВР} = 2,5 * 10^{-5}$ |

Таким образом, исходя из рассчитанных данных, интенсивность отказов технического оборудования ИС НКУ λ_1 равна:

$$\lambda_1 = \lambda_{УЭВМ} + \lambda_{ППЭВМ} + \lambda_{ПИВР} + \lambda_{ППАКТК} + \lambda_{СИ} = 2,34 * 10^{-4}, \left[\frac{1}{ч} \right], \quad (23)$$

По полученным данным в приведённом примере реализации разработанной математической модели по формуле (12) был построен график, отражающий надёжность информационной структуры наземного комплекса управления роём БПЛА (рис.5).

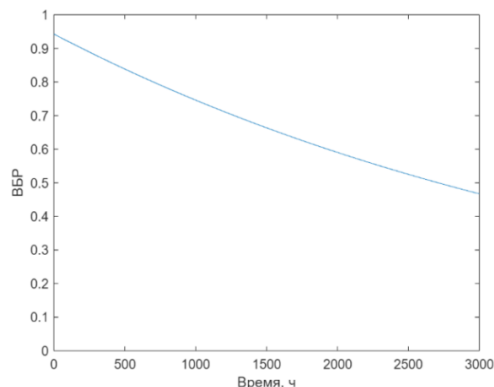


Рис.5. График зависимости ВБР ИС НКУ от времени оперативной работы по разработанной модели

Заключение

Проведён анализ методов и математических моделей оценки надёжности информационных систем применительно к структуре ТС НКУ и ПО, необходимых для управления роём беспилотных летательных аппаратов.

На базе этого анализа обосновано использование следующих математических моделей:

1. Для оценки надёжности ТС НКУ целесообразно использовать метод экспоненциального закона надёжности.

2. Для оценки надёжности программного обеспечения целесообразно использовать модель Джелински-Моранды.

Для оценки надёжности информационной структуры НКУ роём БПЛА разработана математическая модель, которая представляет собой композицию этих моделей.

Приведена реализация разработанной математической модели на конкретном примере.

Список литературы

1. Баранова, А.В., Ямпурин, Н.П., Методы оценки надёжности информационных систем. – URL: <https://cyberleninka.ru/article/n/metody-otsenki-nadezhnosti-informatsionnyh-sistem/viewer> (дата обращения 2023-23-11).

2. Макаров, И.В., Кокорин, В.И., Комплекс управления беспилотными летательными аппаратами для дистанционного зондирования земли. – URL: <https://uav-siberia.com/news/kompleks-upravleniya-bes-pilotnymi-letatelnyimi-apparatami-dlya-distantsionnogo-zondirovaniya-zemli/> (дата обращения 2023-20-12).

3. Чекал, Е.Г., Чичев, А.А., Надёжность информационных систем: учебное пособие. – Ульяновск : УлГУ, 2012. – 118 с.

РЕАЛИЗАЦИЯ ЭЛЕКТРО-ТЕПЛОВЫХ РАСЧЕТОВ СИЛОВЫХ СХЕМ В PYTHON SPICE С ИСПОЛЬЗОВАНИЕМ ЦИФРОВЫХ ДВОЙНИКОВ МОЩНЫХ МОП ТРАНЗИСТОРОВ

Зубкова А.И., Харитонов И.А.

Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Работа посвящена обработке процесса электро-теплого моделирования работы мощных схем в среде Python Spice на примере МОП транзистора, с использованием цифрового двойника этого МОПТ. На основе экспериментальных данных отработаны и включены в модель МОПТ зависимости его порогового напряжения и крутизны от температуры. Проведено моделирование ВАХ транзистора и верификация его электро-тепловой модели по результатам измерений его электро-тепловых характеристик с учетом эффектов разогрева.

Введение

Процесс моделирования схем является относительно сложной задачей из-за наличия большого количества компонентов, их режимов работы, разнообразия способов реализации одной и той же физической задачи. Для изучения поведения электрической цепи необходимо выполнить несколько итераций симуляций в нужной среде и при взаимодействии схем и подсхем количество итераций увеличивается, значит увеличивается время моделирования. Поэтому существует необходимость в ускорении процесса и необходимость в автоматизированной аналитике современных электронных схем, в связи с их сложными характеристиками. При этом возможности компьютерного моделирования улучшаются, в связи с добавлением новых библиотек, искусственного интеллекта и машинного обучения.

Благодаря функциональности Python Spice можно передавать параметры схемы, создавать сценарии моделирования, управлять выходными данными моделирования Spice, анализировать и обрабатывать выходные данные [1]. В этом качестве Python имеет следующие преимущества при SPICE анализе схем:

1. совмещение с объектно-ориентированным программированием (ООП) [2];
2. возможность создания многоцветных и модульных компонентов, которые можно параметризовать и создать общий шаблон для моделирования различных схем;
3. неограниченность графическим интерфейсом при создании электрических схем и использование скриптов для быстрого создания сложных схем;
4. неизменность языка программирования, в то время как графический интерфейс изменяется и приходится пересобирать схемы;
5. возможность встраивать параметры в сами компоненты и проводить моделирование одновре-

менно при разных зависимостях, в разных средах и не проводит последовательное моделирование, разделяя на под схемы;

6. возможность использовать цифровые двойники, которые разработаны на языке Python.

Цифровые двойники интенсивно развиваются во всем мире. Цифровой двойник – это мост между физическим и виртуальным объектом. ЦД позволяют подробно исследовать поведение объекта, не прибегая к использованию реального объекта [3].

Цифровой двойник включает в себя:

1. Математическая модель объекта.
2. Развивающийся набор данных, относящихся к объекту.
3. Средство динамического обновления или корректировки модели в соответствии с данными[4].

Анализ литературы показал, что Python Spice эффективен и для реализации электро-тепловых расчетов силовых схем с использованием цифровых двойников [3, 6].

Таким образом, Python облегчает и делает эффективнее spice моделирование, так как позволяет совмещать использование различных библиотек и высокоуровневое программирование.

Учет зависимости порогового напряжения от температуры при моделировании мощного МОП транзистора

Для проведения моделирования был разработан и затем использован цифровой двойник мощного МОП транзистора IRF610[5].

По предоставленным экспериментальным данным построена зависимость порогового напряжения от температуры в диапазоне от 0 до 120° в среде Excel, рис.1.

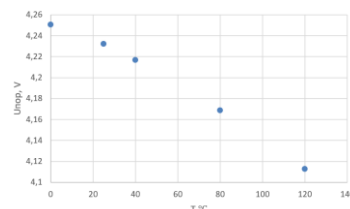


Рис.1. Измеренная зависимость порогового напряжения от температуры

На основании анализа зависимости, с помощью метода аппроксимации была подобрана формула (1) зависимости корректирующей добавки для порогового напряжения от температуры. В связи с анализом литературы по температурным зависимостям параметров МОП транзисторов, было принято решение задавать формулу в виде линейной зависимости. Подобранная формула дает погрешность в пределах 2%.

$$V_{корр} = 0.008 - 0.00115 * (T-25) \quad (1)$$

Моделирование и построение графиков зависимости тока стока от напряжения затвора было проведено с добавлением в схему зависимого источника напряжения на затворе, с помощью среды Ltspice и Python Spice.

Учет зависимости крутизны от температуры при электрическом моделировании мощного МОП транзистора

Следующим этапом, по данным эксперимента, с помощью метода аппроксимации, была подобрана формула зависимости крутизны от температуры (2) и для моделирования ВАХ МОП был добавлен в схему корректирующий источник тока.

$$I_{\text{корр}} = \left(1 - \frac{T+273}{25+273}^{-1.5}\right) * I(V4) \quad (2)$$

Для проверки результаты моделирования сравниваются с результатами эксперимента, построенными на рис.2.

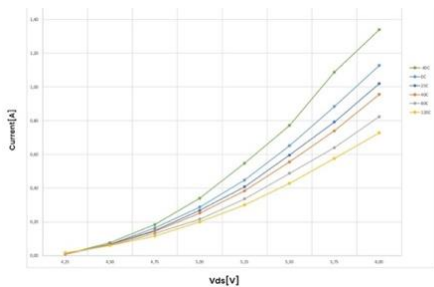


Рис.2. Измеренные (точки) и смоделированные зависимости I_c после коррекции для $U_{\text{пор}}$ (линии)

Моделирование и построение графиков зависимости I после коррекции от U_z было проведено с помощью среды Python Spice, рис.3.

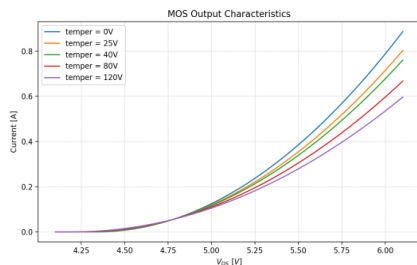


Рис.3. Зависимость I после коррекции от U_z , моделирование с помощью Python spice

При увеличении температуры окружающей среды снижаются I_c , $R_{\text{пор}}$ и коэффициент наклона сток-затворной характеристики. По результатам моделирования и эксперимента можно говорить о правильном подборе основных формул (1) и (2).

Электро-тепловое моделирование мощного МОП транзистора

Вначале было проведено моделирование без саморазогрева транзистора. Результаты измерений выходных ВАХ (в динамическом режиме) транзистора представлены на рис.4.

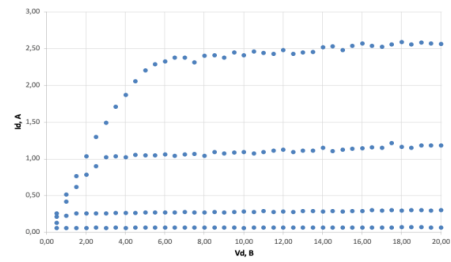


Рис.4. Зависимость тока стока от напряжения стока, при $U_g = 4.5, 5, 6, 7$, по данным измерений при комнатной температуре, 25°C

Результаты моделирование приведены: в среде Python spice – рис.5.

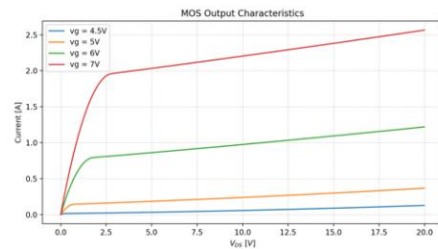


Рис.5. Зависимость тока стока от напряжения стока, при $U_g = 4.5, 5, 6, 7$ при комнатной температуре, 25°C , моделирование в среде Python spice

На следующем этапе были проведены измерения температурных характеристик транзистора IRF610 с 4 видами теплоотводов, представленными в таблице 1.

Таблица 1. Характеристики радиаторов

| N | Вид р-ра | m, г | Спов., см ² | T кр., C |
|---|------------|------|------------------------|----------|
| 1 | u-образный | 1,5 | 18 | 120 |
| 2 | u-образный | 3,1 | 30 | 86 |
| 3 | ребристый | 100 | 280 | 49 |
| 4 | ребристый | 1000 | 2000 | 32 |

Измерения проводилась при подаче постоянного напряжения 6В, на затвор; 4В, на сток. Температура измерялась при помощи тепловизора.

Результаты измерений саморазогрева транзистора IRF610 с несколькими видами радиаторов приведены на рис.6.

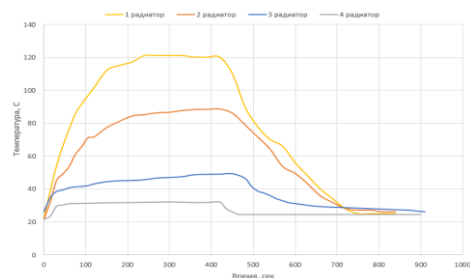


Рис.6. Изменение температуры транзистора с течением времени для 4 видов радиаторов

Проведено температурное моделирование для каждого из радиаторов в среде Python spice, рис.7.

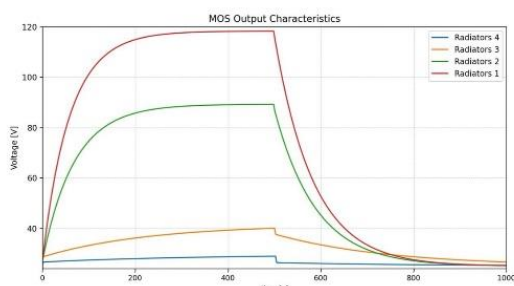


Рис.7. Моделирование изменения температуры транзистора с течением времени для 4 видов радиаторов с помощью Python spice

Последние два радиатора подходят для схем с большим энергопотреблением и нагреванием. В то время как радиаторы 1,2 для более меньших значений, так как отводят меньше тепла.

Для транзистора IRF610 был выбран радиатор 2 и к основной схеме была добавлена тепловая подсхема и проведено моделирование в среде Python spice, рис.8.

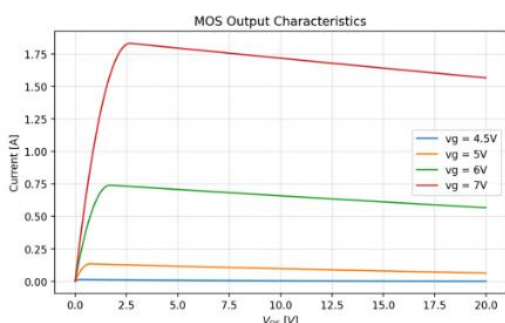


Рис.8. Зависимость тока стока от напряжения стока, при $U_g = 4.5, 5, 6, 7$ при комнатной температуре, 25°C , с тепловой схемой, моделирование в среде Python spice

По построенным графикам можно видеть, что при добавлении тепловой схемы, при высоких напряжениях на стоке наблюдается спад тока вследствие саморазогрева МОП транзистора. В связи с увеличением напряжения, увеличивается температура в канале, что приводит к уменьшению тока стока и ухудшению характеристик транзистора [9].

Заключение

Проведены измерения саморазогрева мощного МОП транзистора для нескольких видов радиаторов.

По результатам измеренных данных были получены и включены в модель и цифровой двойник мощного МОП транзистора зависимости порогового напряжения и крутизны от температуры. Погрешность моделирования (основана на сравнении точек экспериментальных данных и моделирования в среде Python spice) лежит в пределах 5%.

Проведено электро-тепловое моделирование в среде Python spice с использованием цифрового двойника мощного МОП транзистора, с настройкой моделей транзистора и тепловой подсхемы по экспериментальным данным для нескольких видов радиаторов.

Предполагается, что результаты работы будут в дальнейшем использованы при проведении электро-теплого моделирования схем с использованием среды Python Spice с другими схемами и цифровыми двойниками.

Список литературы

1. Digital Machine. Basics of circuit simulation with python spice – URL: <https://thedigitalmachine.net/basics-of-circuit-simulation-with-python-spice/> (дата обращения: 21.01.2024).
2. Olzhas S. Tazabekov. Digital to Analog Converter Modelling with PySpice (2017) – URL: <http://www.tazabekov.com/blog/2017/06/digital-to-analog-converter-modelling-with-pyspice/> (дата обращения: 01.02.2024).
3. Sergio de López Diz. A real-time digital twin approach on three-phase power converters applied to condition monitoring // Applied Energy 334 (2023) 120606
4. Зубкова А.И. Харитонов И.А. Разработка цифровых двойников для мощных МОП транзисторов// Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов имени Е.В.Арменского – 2023.
5. Vishay siliconix. IRF610 Datasheet – URL: <https://www.vishay.com/docs/91023/irf610.pdf> (дата обращения: 12.02.2024).
6. Leander Van Cappellen. A Real-time Physics Based Digital Twin for Online MOSFET Condition Monitoring in PV Converter Applications // 2022 28th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC)
7. Зубкова А.И. Реализация цифровых двойников для мощных МОП транзисторов с помощью ПО Python// Российский форум «Микроэлектроника 2024», Школа молодых ученых.
8. Хоровиц П., Хилл У. Искусство схемотехники //М: Мир. – 1998.
9. Масальский Н.В. Влияние эффекта самонагрева на вольт-амперные характеристики полевых транзисторов на структуре «кремний на изоляторе» при высоких температурах// Радиотехника и электроника – 2019.

ЭКСПЕРИМЕНТАЛЬНОЕ СРАВНЕНИЕ МЕТОДОВ СИНТЕЗА ЛОГИЧЕСКИХ СХЕМ

Ягжов А.А.¹, Вершков М.Д.², Романов Н.С.²,
Знатнов Е.П.³, Федотова А.А.³

Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент электронной инженерии,

²департамент прикладной математики
МИЭМ НИУ ВШЭ,

³департамент программной инженерии
ФКН НИУ ВШЭ

Аннотация

В данной работе представлены результаты экспериментального сравнения методов синтеза комбинационных логических схем, реализующих заданные булевы функции. Сравнение основывалось на оценках энергопотребления, глубины и площади синтезируемых логических схем. Рассмотрены были следующие методы: Акерса, би-декомпозиции, каскадов, Минато-Морреале, Рида-Маллера и DSD - разложения.

Введение

Технологически независимая оптимизация (оптимизация) логических схем является одним из ключевых этапов логического синтеза и во многом определяет качество проектируемой аппаратуры.

Различные методы синтеза логических схем, реализующих заданные булевы функции (методы синтеза) находят широкое применение в оптимизации. Например, данные методы могут использоваться в подходах, основанных на переписывании (rewriting), которое заключается в замене подсхем исходной логической схемы на другие подсхемы, более эффективно реализующие те же функции [1].

Оптимизация может вестись по трем основным критериям (в скобках указаны метрики оценки на уровне логических схем):

- 1) энергопотребление (переключательная активность);
- 2) глубина (длина максимального пути от входов к выходам);
- 3) площадь (число логических вентилях).

Целью описанного в данной работе исследования являлось проведение экспериментального сравнения различных методов синтеза и определение лучших из них для каждого критерия оптимизации.

Результаты сравнения различных методов синтеза встречаются и в других работах. Например, в статье [2] описаны результаты сравнения таких методов, как: синтез схем по ДНФ и КНФ исходной булевой функции, использование мультиплексорно-ориентированных выражений (MUX-based expressions), метод Квайна-Мак-Класки, синтез схем по полиномам (XOR expressions). Сравнение методов синтеза основывалось на оценках площади и глубины синтезируемых схем, а число аргументов исходных булевых функций не превышало 5.

В работе [3], было проведено сравнение количества элементарных конъюнкций, содержащихся в минимальных ДНФ булевых функций, с количеством

элементарных конъюнкций в ДНФ, полученных с помощью метода Минато-Морреале [4]. Сравнение было проведено для различных булевых функций с числом аргументов, не превышающим 20.

В данной статье описаны результаты сравнения методов синтеза, в том числе не рассмотренных в приведенных примерах. Сравнение было основано на оценках не только глубины и площади синтезируемых схем, но и на оценке энергопотребления. Эксперимент был проведен на булевых функциях от 3 до 10 аргументов включительно. Отличительной особенностью данного эксперимента является использование ограничения на максимальное число входов логических вентилях, которое в данном эксперименте равнялось трем.

В следующих разделах приведены краткие описания каждого из рассмотренных методов, методика и результаты проведенных экспериментов, итоги исследования.

Обзор методов

Метод Акерса [5] предназначен для построения схемы в базисе из следующих элементов: мажоритарная функция, НЕ, константа 0. Алгоритм является итерационным и основан на преобразованиях таблицы, составленной на базе таблицы истинности исходной функции. Колонки таблицы, которая преобразовывается, выступают в качестве аргументов для вентилях. На каждой итерации перебираются все тройки колонок, от которых вычисляется таблица истинности для мажоритарного вентиля. Полученные таблицы истинности сравниваются между собой, затем выбирается лучшая в соответствии с некоторыми эвристиками. Далее она заносится в таблицу для преобразований в качестве еще одной колонки. Алгоритм заканчивается при получении таблицы истинности, равной таблице истинности исходной функции.

Метод би-декомпозиции [6] основан на разложении исходной булевой функции на две функции от меньшего числа аргументов, которые связываются одной из следующих операций: И (И-НЕ), ИЛИ (ИЛИ-НЕ), исключающее ИЛИ. Полученные функции рекурсивно декомпозируются аналогичным образом. В рамках данной работы был рассмотрен метод эвристической би-декомпозиции, когда аргументы новых функций вычисляются в процессе разложения исходной.

Метод каскадов [7] основан на рекурсивном разложении функции с использованием разложения Шеннона [8]:

$$f(x_1, x_2, \dots, x_n) = x_n \cdot f(x_1, x_2, \dots, x_{n-1}, 1) \cup \bar{x}_n \cdot f(x_1, x_2, \dots, x_{n-1}, 0), \quad (1)$$

где f – булева функция, x_i – i -й аргумент функции.

На каждом этапе разложения производится упрощение логической функции с помощью алгебраических преобразований, если оно возможно.

Метод Минато-Морреале [4] основан на получении безыбыточной ДНФ, из которой нельзя удалить ни одной конъюнкции и ни одного литерала из конъюнкции

юнкции так, чтобы она оставалась тождественной исходной булевой функции. Данный метод не гарантирует, что полученная ДНФ будет минимальной.

Метод Рида-Маллера [9] предполагает получение полинома Рида-Маллера, который представляет собой элементарные конъюнкции, связанные суммой по модулю 2. В элементарные конъюнкции одна часть переменных входит только с отрицанием, а другая – только без отрицания. В данном исследовании рассматривался частный случай полинома Рида-Маллера, в который все переменные входят без отрицания (полином Жегалкина [10]).

Метод DSD-разложения (Disjoint Support Decomposition) [11] основан на разложении исходной булевой функции на несколько подфункций от меньшего числа аргументов. Наборы аргументов для каждой полученной подфункции не должны пересекаться между собой. Количество подфункций больше либо равно двум и зависит от исходной функции. Подфункции могут быть связаны ассоциативным оператором (И, ИЛИ, исключающее ИЛИ) либо произвольной функцией от трех и более аргументов, которая не имеет DSD-разложения. Полученные подфункции рекурсивно декомпозируются аналогичным образом.

В таблице 1 представлены логические элементы, в базисе которых синтезировались схемы каждым из методов синтеза.

Таблица 1. Базис синтезируемых схем

| Метод синтеза | Логические элементы |
|-----------------------|---|
| Метод Акерса | мажоритарная функция, НЕ, константа 0 |
| Метод би-декомпозиции | И, НЕ |
| Метод каскадов | И, НЕ, константы 1 и 0 |
| Метод Минато-Морреале | И, НЕ. |
| Метод DSD-разложения | И, НЕ, исключающее ИЛИ, константы 1 и 0 |
| Метод Рида-Маллера | И, исключающее ИЛИ, константа 1. |

Методика испытаний

Методы синтеза были реализованы на языке C++. Для методов Минато-Морреале и DSD-разложения использовались библиотеки kitty [12] и STACCATO [13] соответственно. Для сравнения методов между собой был написан набор тестовых примеров, в котором равновероятно генерировалось определенное количество таблиц истинности булевых функций от трех до десяти аргументов с помощью инструментов, предоставляемых библиотекой kitty [12]. Количество генерируемых таблиц истинности булевых функций от конкретного числа аргументов равнялось 2^n , где n – число аргументов.

Данные таблицы истинности подавались в качестве входных параметров методам синтеза, которые строили схемы, состоящие из логических вентилях, арность которых не превышала трех. Для сравнения схем использовались следующие характеристики:

количество аргументов функции, площадь схемы (количество вентилях), глубина (длина максимального пути в схеме), переключательная активность. Переключательная активность схемы рассчитывалась как сумма вероятностей переключения всех ее вентилях:

$$Z = \sum_{i=0}^n \frac{s_i}{t}, \quad (2)$$

где Z – переключательная активность логической схемы, s_i – число переключений ячейки (из 1 в 0 и из 0 в 1), i – индекс ячейки логической схемы, t – число симуляций.

При проведении эксперимента число симуляций каждой логической схемы составляло 1024.

Реализации методов Акерса и би-декомпозиции проверялись на таблицах истинности булевых функций от 3-7 и 3-8 аргументов соответственно. Причина состоит в высокой асимптотической сложности данных методов [14, 6].

Результаты

Результаты сравнения методов синтеза на описанном выше тестовом наборе представлены в таблице 2. В ее ячейках содержатся средние арифметические значения полученных результатов.

Часть таблицы 2, соответствующая переключательной активности схем, демонстрирует, что лидером по оптимизации данного параметра для функций от четырех до десяти аргументов включительно является метод Минато-Морреале, для функций от трех аргументов – метод Акерса. При этом, метод би-декомпозиции является вторым по оптимизации для функций от пяти до восьми аргументов включительно. Третьим – метод Акерса, четвертым – метод Рида-Маллера, методы каскадов и DSD-разложения – пятый и шестой соответственно.

Согласно таблице 2, методы би-декомпозиции и Минато-Морреале лучше остальных методов оптимизируют глубину схемы (би-декомпозиции – для функций от четырех, пяти и восьми аргументов, Минато-Морреале – для остальных). За ними следует метод Рида-Маллера, затем Акерса, DSD-разложения и каскадов.

Часть таблицы 2 «Число логических вентилях» показывает, что метод Акерса предпочтительнее остальных рассматриваемых методов для оптимизации площади для функций от трех до семи аргументов включительно, для остальных функций – метод Минато-Морреале. Третьим в части оптимизации количества вентилях является метод би-декомпозиции, четвертым – DSD-разложения, пятым – каскадов, замыкает данный рейтинг метод Рида-Маллера.

Результаты эксперимента были сопоставлены с работами [2] и [3]. Размеры получаемых безызбыточных ДНФ были сопоставимы с результатами статьи [3]. Метод Рида-Маллера показал в данной работе более оптимальные характеристики синтезируемых схем, что было вызвано использованием логических вентилях с числом входов равным трем. В статье же [2] рассматривались только вентиля с двумя входами.

Таблица 2. Результаты экспериментов

| Количество аргументов функции | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------------------|-------------|-------------|--------------|--------------|--------------|---------------|---------------|---------------|
| Переключательная активность схемы | | | | | | | | |
| Метод Акерса | 3 | 4.8 | 7.59 | 14.35 | 29.63 | | | |
| Метод би-декомпозиции | 3.2 | 4.98 | 6.91 | 10.53 | 16.55 | 31.82 | | |
| Методы каскадов | 3.26 | 6.74 | 12.93 | 24.93 | 46.79 | 92.43 | 181.67 | 361.69 |
| Метод Минато-Морреале | 3.14 | 4.42 | 6.47 | 9.87 | 16.34 | 31.38 | 55.8 | 112.57 |
| Метод DSD-разложения | 3.35 | 6.7 | 13.07 | 24.99 | 47.64 | 93.66 | 184.39 | 366.11 |
| Метод Рида-Маллера | 4.38 | 6.75 | 11.55 | 19.03 | 33.85 | 61.01 | 116.16 | 227.7 |
| Глубина схемы | | | | | | | | |
| Метод Акерса | 2.38 | 3.94 | 6 | 9.73 | 18.84 | | | |
| Метод би-декомпозиции | 2 | 3.13 | 4.22 | 5.03 | 5.71 | 6.33 | | |
| Методы каскадов | 2.75 | 5.63 | 7.88 | 10 | 12 | 14 | 16 | 18 |
| Метод Минато-Морреале | 1.88 | 3.44 | 4.31 | 5 | 5.59 | 6.48 | 7.99 | 8.51 |
| Метод DSD-разложения | 2.5 | 5.25 | 7.47 | 9.84 | 11.98 | 14 | 16 | 18 |
| Метод Рида-Маллера | 2.63 | 3.56 | 4.78 | 5.64 | 6.39 | 7 | 8.19 | 9 |
| Число логических вентилей | | | | | | | | |
| Метод Акерса | 3.13 | 7.13 | 12.22 | 25.47 | 56.72 | | | |
| Метод би-декомпозиции | 3.25 | 7.38 | 16.09 | 37.05 | 77.31 | 178.44 | | |
| Методы каскадов | 3.63 | 10.88 | 25.06 | 53.97 | 108.78 | 222.52 | 446.91 | 899.74 |
| Метод Минато-Морреале | 3.13 | 7.13 | 16.31 | 35.79 | 77.71 | 177.41 | 371.58 | 836.1 |
| Метод DSD-разложения | 3.5 | 10.69 | 24.78 | 53.39 | 108.26 | 221.11 | 445.25 | 895.06 |
| Метод Рида-Маллера | 6.38 | 12.13 | 27.94 | 59.4 | 133.65 | 292.18 | 643.41 | 1411.45 |

Заключение

Были проанализированы результаты работы шести методов синтеза логических схем, реализующих заданные булевы функции. Сравнение полученных схем проводилось по трем основным критериям: число вентилей, глубина и переключательная активность. Анализ результатов показал, что для оптимизации площади схемы для функций от трех до семи аргументов включительно предпочтителен выбор метода Акерса, для иного числа аргументов – метод Минато-Морреале. В оптимизации глубины схемы лидируют методы би-декомпозиции (для функций от четырех, пяти и восьми аргументов) и Минато-Морреале (для остального числа аргументов). Для оптимизации переключательной активности лучше остальных рассматриваемых методов проявил себя метод Минато-Морреале.

Список литературы

1. Riener, H. Exact DAG-aware rewriting / H. Riener, A. Mishchenko, M. Soeken. – Текст : непосредственный // 2020 Design, Automation & Test in Europe Conference & Exhibition (DATE). – IEEE, 2020. – С. 732-737.
2. Ammes, G. A. Comparative Analysis of Different Boolean Function Synthesis Methods / G. Ammes, W. Lau, R. P. Ribas. – Текст : электронный // SBMicro : [сайт]. – 2018. – URL: <https://sbmicro.org.br/eventos/sforum/volume-18> (дата обращения: 27.01.2024).
3. Sasao T. Worst and best irredundant sum-of-products expressions / T. Sasao, J. T. Butler. – Текст : непосредственный // IEEE Transactions on Computers. – 2001. – С. 935-948.
4. Minato, S. Fast generation of prime-irredundant covers from binary decision diagrams / S. Minato. –

Текст : непосредственный // IEICE transactions on fundamentals of electronics, communications and computer sciences. – 1993. – Т. 76. – №. 6. – С. 967-973.

5. Akers, S. B. Synthesis of combinational logic using three-input majority gates / S. B. Akers. – Текст : непосредственный // 3rd annual symposium on switching circuit theory and logical design (SWCT 1962). – IEEE, 1962. – С. 149-158.

6. Pottosin, Y. V. Synthesis of combinational circuits by means of bi-decomposition of Boolean functions / Y. V. Pottosin. – Текст : непосредственный // Прикладная дискретная математика. – 2023. – №. 60. – С. 95-105.

7. Поваров, Г. Н. Метод синтеза вычислительных и управляющих контактных схем / Г. Н. Поваров. – Текст : непосредственный // Автоматика и телемеханика. – 1957. – Т. 18. – №. 2. – С. 145-162.

8. Shannon, C. E. The synthesis of two-terminal switching circuits / C. E. Shannon. – Текст : электронный // The Bell System Technical Journal. – 1949. – Т. 28. – №. 1. – С. 59-98. – URL: https://ieeexplore.ieee.org/iel7/6731005/6771694/06771698.pdf?casa_token=XpXJek7KvK4AAAAA:1UDentJWVQNA6rexzzGiUySq_wSdnEiuLEuro5LXj3OV7AIO9Zra5va5w8VHqkJ-1y6vA0iaRUIZ (дата обращения: 27.01.2024)

9. Harking, B. Efficient algorithm for canonical Reed-Muller expansions of Boolean functions / B. Harking. – Текст : непосредственный // IEE Proceedings E (Computers and Digital Techniques). – 1990. – Т. 137. – №. 5. – С. 366-370.

10. Капитонова, Ю. В. Лекции по дискретной математике : учебник / Ю. В. Капитонова, С. Л. Кривой, А. А. Летичевский. – Санкт-Петербург

: БХВ-Петербург, 2004. – 624 с – ISBN 5-94157-546-7. – Текст : непосредственный.

11. Bertacco, V. Scalable Hardware Verification with Symbolic Simulation : монография / V. Bertacco. – New York : Springer New York, 2006. – 180 с. – ISBN 978-0-387-29906-8. – Текст : непосредственный.

12. Репозиторий kitty : [сайт]. – 2024. – URL: <https://github.com/msoeken/kitty> (дата обращения: 27.01.2024). – Текст : электронный.

13. Библиотека STACCATO : [сайт]. – 2024. – URL: <https://web.eecs.umich.edu/staccato/> (дата обращения: 27.01.2024). – Текст : электронный.

14. Lee, S. Y. Riener H., De Micheli G. Logic resynthesis of majority-based circuits by top-down decomposition / S. Y. Lee, H. Riener, G. De Micheli. – Текст : непосредственный // 2021 24th International Symposium on Design and Diagnostics of Electronic Circuits & Systems (DDECS). – IEEE, 2021. – С. 105-110.

ПРОГРАММНО-АППАРАТНАЯ РЕАЛИЗАЦИЯ МЕТОДИКИ УСКОРЕННЫХ ИСПЫТАНИЙ МАЛОМОЩНЫХ РЕЗИСТОРОВ

*Купцова Д.Д., Кучина А.П., Старкова Э.А.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Данная работа содержит описание контрольно-измерительного модуля и интерфейса пользователя для проведения ускоренных испытаний резисторов и верификации их показателей надежности. Разработанный комплекс позволяет определять показатели надежности измеряемых компонентов - резисторов. Далее после обработки измеренных данных определяется интенсивность отказа компонентов. Благодаря используемой методике ускоренных испытаний можно в короткие сроки получить показатели надежности резисторов. Результатом работы является разработанная схема электрическая принципиальная контрольно-измерительного модуля и интерфейс пользователя.

Введение

В настоящее время с развитием рынка радиоэлектронных изделий важным фактором их конкурентоспособности является надежность. Изучение холодных земель, арктических областей, космоса является объектом интереса человечества так же, как и применение устройств в экстремальных условиях, как климатических, таких как низкие (до - 55 °С) и высокие (до +125 °С) температуры, так и электрических - мощные электрические нагрузки, скачки напряжения, и механических - воздействие на изделие ударов и механических вибраций. Для этого требуется аппаратура, способная работать в температурных услови-

ях, выходящих за рамки среднестатистических диапазонов. Важно обеспечить высокую надежность радиоэлектронных изделий, входящих в состав устройств, применяемых в экстремальных условиях, для снижения количества их отказов и увеличения конкурентоспособности на мировом рынке отечественных производителей электроники. Это может быть достигнуто путем определения показателей надежности в большом диапазоне температур.

Проведение ускоренных испытаний в соответствии с разработанной методикой обусловлено необходимостью проверки полученных математических выкладок об увеличении эксплуатационной интенсивности отказов в зоне отрицательных температур, что позволит использовать уточненную математическую модель, учитывающую деградиционные процессы, протекающие в резисторах. В качестве новизны выступают статистические данные об интенсивности отказов резисторов на расширенном температурном диапазоне от -55 °С до +125 °С.

Целью данной работы является реализация методики ускоренных испытаний резисторов посредством разработки контрольно-измерительного модуля (КИМ) параметров исследуемых элементов для определения их эксплуатационной интенсивности отказов.

Данная статья содержит три раздела. В первом разделе описана используемая методика проведения испытаний. Во втором разделе приведено описание контрольно-измерительного модуля. Третий раздел содержит в себе описание программной части КИМ.

Описание методики ускоренных испытаний

В рамках данной работы была разработана методика ускоренных испытаний [2], направленная на облегчение проведения ускоренного тестирования надежности резисторов, отнесенных к аппаратуре группы 1.1 в соответствии со стандартами ГОСТ РВ 20.39.304 [3].

Структура программно-аппаратного комплекса представлена на рис.1.

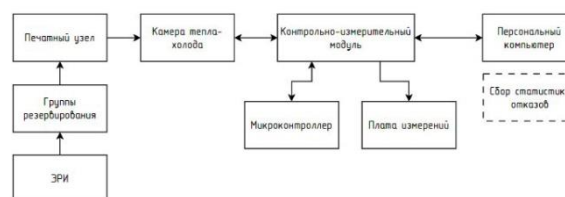


Рис.1. Структура программно-аппаратного комплекса

Печатный узел состоит из ЭРИ (резисторов), обрезающих группы резервирования, которые являются объектом исследования. Климатическая испытательная камера, в которую помещается исследуемый печатный узел, предназначена для проведения ускоренных испытаний. КИМ состоит из микроконтроллера и платы измерений. На ПК размещен интерфейс пользователя для сбора данных с определенных

посредством программного инструмента контрольных точек и управления.

Контролируя температуру тестирования с помощью камеры тепла-холода (КТХ), повышая и понижая её по сравнению с рабочей температурой +25°C, а также увеличивая частоту циклов включения/выключения напряжения питания печатной платы, можно сократить время тестирования и смоделировать реальные условия за более короткое время. Такой подход позволяет контролировать такие показатели надежности, как вероятность безотказной работы, средняя наработка на отказ и интенсивность отказов.

Циклограмма отражает последовательность и длительность операций, входящих в цикл работы КТХ (рис.2). Разработанная циклограмма для проведения форсированных испытаний состоит из четырех основных этапов: испытания в нормальных климатических условиях (длительность 1 этапа - 20 часов), воздействие пониженной температуры (длительность 2 этапа – 10 часов), циклические колебания температуры (длительность 3 этапа – 10 часов) и воздействие повышенной температуры (длительность 4 этапа – 30 часов). Задачей этого комплексного режима испытаний является проверка надежности резисторов в течение запланированного срока службы 43800 часов.

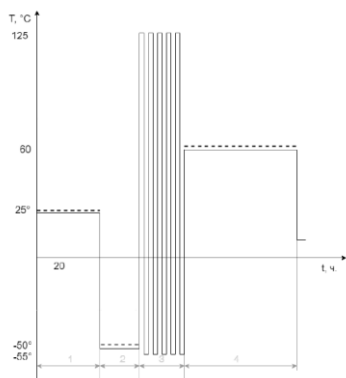


Рис.2. Циклограмма ускоренных испытаний
(- - - - повышенная частота циклов
«включено-выключено» для питания)

Соединение с КТХ реализовано по протоколу Modbus TCP при помощи библиотеки EasyModbus, что позволяет регулировать температуру в КТХ дистанционно в соответствии с определенным этапом испытаний. Управление коэффициентом нагрузки и циклами включения/выключения реализовано программно в соответствии с методикой.

Структура контрольно-измерительного модуля

Контрольно-измерительный модуль представляет собой совокупность двух печатных плат с электрорадиоизделиями: на первой плате размещаются пассивные элементы (резисторы), а вторая плата осуществляет измерение параметров резисторов, таких как ток и напряжение. Схема электрическая принципиальная второй части контрольно-измерительного

модуля изображена на рис.3. Мультиплексоры MULT1, MULT2 (74hc4051) коммутируют входы в один единственный выход, передают сигналы на пины микроконтроллера, установленного в платформе Arduino. После обработки пришедших данных получаем информацию о значениях потенциала на двух концах резистора, из чего в дальнейшем узнаем напряжение на резисторе. Мультиплексор MULT3 позволяет обработать данные, полученные с датчиков тока, работающих на эффекте Холла (U1, U2, U3, U4, U5, U6, U7, U8) и также отправляет на Arduino. То есть данный КИМ позволяет получить значения напряжения на двух концах резисторов и значения тока, протекающего через эти резисторы. Данная система реализована в виде двух печатных плат. Первая плата содержит в себе резисторы и разъемы для выводных проводов (плата групп резервирования). На данном этапе разработки комплекса предполагается использование одной группы резисторов единоразово. Для последующих тестов необходимо заново припаять новую группу резисторов. Эта плата помещается в тепловую камеру для дальнейшего исследования. Вторая плата (плата управления) содержит в себе датчики тока, мультиплексоры, системы питания, а также разъемы для взаимодействия двух плат. Плата управления не подвергается воздействию температур, а только осуществляет измерение параметров изучаемых элементов - резисторов и управляет напряжением, подаваемым на резисторы.

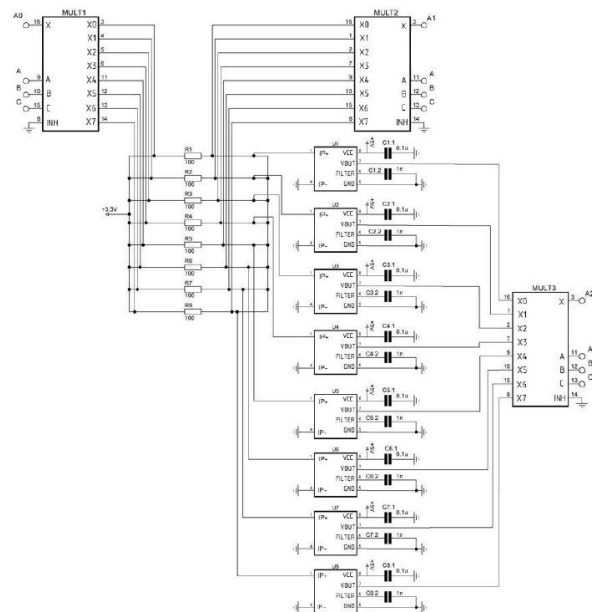


Рис.3. Схема электрическая принципиальная
контрольно-измерительного модуля

Данная схема не содержит никаких внешних защитных систем, так как платформа Arduino предусматривает защиту от короткого замыкания, перенапряжения.

На печатные платы подается напряжение 5В. Питание подается от однофазной трёхпроводной сети 230В через преобразователь напряжения на 12В и после через понижающий DC-DC преобразователь на 5В.

Описание программной части

В рамках данной работы был разработан пользовательский интерфейс для КИМ (рис.4) с использованием языка программирования “Python” и программы “Qt Designer” [4].

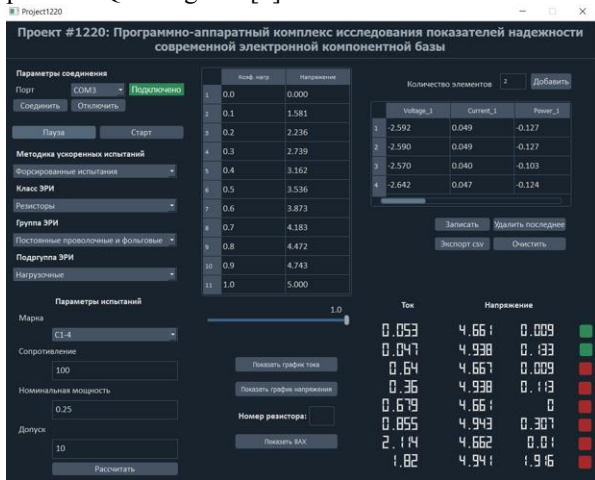


Рис.4. Пользовательский интерфейс

Для того, чтобы начать работать с интерфейсом, необходимо проверить параметры соединения и выбрать нужный COM-порт. Состояние подключения будет отображаться на индикаторе. Данные с платформы Arduino помещаются в буфер памяти, а затем считываются оттуда на ПК. Далее пользователь задает начальные параметры, такие как сопротивление, номинальная мощность и допуск. На основе этих параметров создается таблица коэффициентов нагрузки от 0.1 до 1, позволяющая пользователю вручную выбирать коэффициент нагрузки и соответствующее напряжение, которое будет задано на цифровом пине. Напряжение рассчитывается по формуле (1):

$$U = \sqrt{K_H * R * P_H}, \quad (1)$$

где K_H – коэффициент нагрузки, R – сопротивление исследуемого резистора, P_H – номинальная мощность.

Также в качестве входного параметра пользователем предоставляется количество резисторов, которые в данный момент подключены к мультиплексу. На экране есть индикаторы, которые показывают мгновенные значения потенциалов на концах резистора и ток на каждом резисторе. Ориентируясь на индикаторы, можно добавлять в таблицу строки с мгновенными значениями токов и напряжений. В столбцах содержится напряжение, ток, рассчитанная мощность и фактический номинал для каждого резистора. Сформированную таблицу можно выгрузить в формате .csv для дальнейшей обработки и анализа данных.

Микроконтроллер запрограммирован на считывание значений с обоих концов каждого резистора при помощи двух мультиплексоров для определения напряжения путем АЦП Arduino. Для более точных измерений было учтено опорное напряжение, отно-

сительно которого происходят аналоговые измерения. Третий мультиплексор используется для считывания значений тока на резисторах при помощи подключенных к нему датчиков тока.

Для повышения точности измерений значения с датчика тока усредняются. Передача данных на ПК осуществляется по таймерам для значений тока и напряжений, где формируются строки с заданными ключами, чтобы данные по ключу отображались в нужных ячейках интерфейса.

Также реализована возможность с помощью выбранного коэффициента нагрузки выставлять необходимое напряжение на цифровой пин. Была взята готовая библиотека “PWM.h”, которая позволяет расширить частоту генерации ШИМ-импульсов [5]. Используется Timer1 на 16 бит (вместо стандартных 8 бит). ЦАП происходит в диапазоне от 0 до 65535 (стандартное преобразование от 0 до 1023). Это позволяет повысить точность выставляемого напряжения на пине с точностью до тысячных долей.

При определении фактического номинала резистора были учтены следующие погрешности: относительная погрешность мультиметра, относительная погрешность АЦП/ЦАП преобразования Arduino, относительная погрешность датчика тока. Относительная погрешность измерения определяется по следующей формуле:

$$\epsilon_{\text{изм}} = \sqrt{\epsilon_m^2 + \epsilon_{\text{АЦП}}^2 + \epsilon_d^2}, \quad (2)$$

где ϵ_m^2 – погрешность мультиметра, $\epsilon_{\text{АЦП}}^2$ – погрешность АЦП платформы Arduino, ϵ_d^2 – погрешность датчика тока.

Итоговая погрешность номинала резистора учитывает его допуск и относительную погрешность измерения.

Заключение

В рамках разработанной методики ускоренных испытаний был спроектирован КИМ совместно с пользовательским интерфейсом для проведения форсированных испытаний резисторов с задачей определения показателей надежности.

Проведение ускоренных испытаний в соответствии с разработанной методикой обусловлено необходимостью проверки гипотезы об увеличении интенсивности отказов в зоне отрицательных температур. Впоследствии уточненная математическая модель эксплуатационной интенсивности отказов может быть использована для более точной оценки надежности резисторов.

Результатом работы является разработанная схема электрическая принципиальная контрольно-измерительного модуля и интерфейс пользователя.

Список литературы

1. Абрамов О. В. Актуальность проблемы обеспечения надежности // НиКСС. 2014. №3 (7). URL: <https://cyberleninka.ru/article/n/aktualnost-problemy-obespecheniya-nadezhnosti> (дата обращения: 14.02.2024).

2. Цветков В. Э., Уркунов А. К., Ландер Л. Б., Сафина М. Ф. Разработка структуры программно-аппаратного комплекса для проведения ускоренных испытаний // Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов им. Е.В. Арменского. М.: МИЭМ НИУ ВШЭ, 2023.

3. ГОСТ РВ 20.39.304-98. Комплексная система общих технических требований. Аппаратура, приборы, устройства и оборудование военного назначения.

4. Руководство по Qt Designer // Qt Documentation URL: <https://doc.qt.io/> (дата обращения: 11.02.2024).

5. Увеличение частоты ШИМ // AlexGyver Technologies URL: <https://alexgyver.ru/lessons/pwm-overclock/> (дата обращения: 11.02.2024).

ИССЛЕДОВАНИЕ И МОДЕЛИРОВАНИЕ СТАТИСТИЧЕСКОГО РАЗБРОСА ПАРАМЕТРОВ JFET-ТРАНЗИСТОРОВ С УЧЕТОМ ВЛИЯНИЯ ТЕМПЕРАТУРЫ

Чуча Г.С., Исмаил-Заде М.Р.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

В работе были измерены вольт-амперные характеристики (ВАХ) для набора n - и p -канальных полевых транзисторов с управляющим pn -переходом (JFET) в диапазоне температур от -70 до $+180^{\circ}\text{C}$. На основе результатов измерений проведён статистический анализ основных параметров JFET-транзисторов с учётом влияния температуры и определены параметры SPICE-моделей, учитывающих данное воздействие.

Введение

Транзисторы, используемые в различных электронных устройствах, должны корректно работать при различных условиях. Однако из-за неидеальности технологического процесса возникает разброс параметров, вызванный колебаниями легирующих примесей и шероховатостью кромок затвора. Если не учитывать данный разброс в модели, то это может привести к разбросу параметров интегральной схемы, в которую входят транзисторы. Часть микросхем, параметры которых выходят за границы допустимых значений, уходит в брак [1]. В работах [2] – [4] исследовано влияние температуры на разброс порогового напряжения и крутизны МОП – транзисторов. В данной работе измерения были проведены при большем диапазоне температур. Также был проведён статистический анализ выходной проводимости и коэффициента модуляции длины канала. Данный статистический разброс был учтён в SPICE-модели, что ранее в литературе было отражено недостаточно. Таким образом, целью работы является

исследование и моделирование статистического разброса параметров JFET-транзисторов с учетом воздействия температуры.

Измерение ВАХ транзисторов

В рамках работы были получены сток-затворные и выходные ВАХ у 12 p -канальных и 11 n -канальных JFET-транзисторов в диапазоне температур от -70 до $+180^{\circ}\text{C}$ (от 203 до 453 К). Для измерений использовалась камера тепло-холод, в которую помещались транзисторы, и источник-измеритель Keithley 2602, соединённый с камерой по четырехпроводной схеме подключения.

Статистический анализ параметров транзисторов с учетом влияния температуры

По полученным в ходе измерений ВАХ с помощью программы компьютерного моделирования IC-CAP для каждого транзистора в диапазоне температур от -70 до $+180^{\circ}\text{C}$ были получены следующие параметры: пороговое напряжение v_{to} , крутизна β , выходная проводимость g_{ds} , коэффициент модуляции длины канала λ . Данные параметры являются необходимыми для создания SPICE-модели. Для проведения статистического анализа для каждого параметра были подсчитаны его средние значения, среднеквадратичные отклонения (Sx) и коэффициенты вариации при каждой исследуемой температуре (CV). Таким образом, были получены зависимости среднего значения, Sx и CV от температуры. На рис.1-2 изображены графики зависимости коэффициента вариации порогового напряжения и крутизны от температуры.

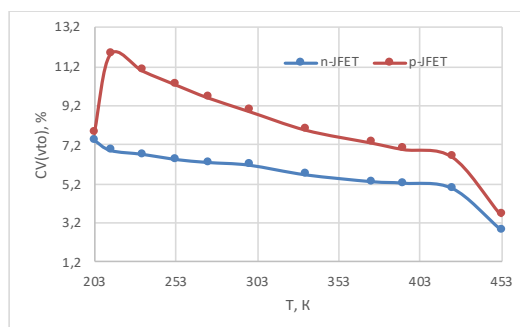


Рис.1. Зависимость коэффициента вариации порогового напряжения v_{to} от температуры

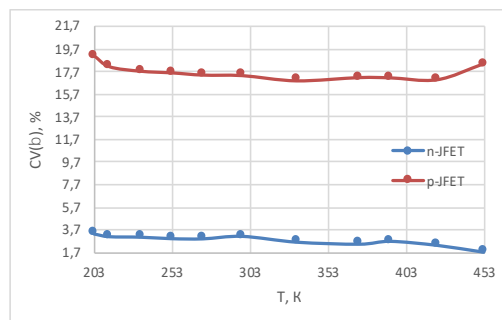


Рис.2. Зависимость коэффициента вариации крутизны β от температуры

Таблица 1. Изменение коэффициента вариации параметров NFET-транзистора при температурах -70 и +180°C относительно комнатной температуры (+25°C), в процентах

| | -70°C | +180°C |
|--|-------|--------|
| Пороговое напряжение v_{to} | +20,8 | -54,1 |
| Крутизна β | +7 | -44,4 |
| Коэффициент модуляции длины канала λ | +9,3 | +17,5 |
| Выходная проводимость g_{ds} | +4,6 | -4,2 |

В приведенной таблице 1 видно, что у n -канального транзистора, по сравнению с комнатной температурой, при повышении температуры до +180°C уменьшаются коэффициенты вариации порогового напряжения, крутизны и выходной проводимости, но увеличивается коэффициент вариации модуляции длины канала. При понижении температуры до -70°C наблюдаем повышение коэффициента вариации всех исследуемых параметров.

Таблица 2. Изменение коэффициента вариации параметров PFET-транзистора при температурах -70 и +180°C относительно комнатной температуры (+25°C), в процентах

| | -70°C | +180°C |
|--|-------|--------|
| Пороговое напряжение v_{to} | +13 | -60 |
| Крутизна β | +9,7 | +5,7 |
| Коэффициент модуляции длины канала λ | +72 | -80,8 |
| Выходная проводимость g_{ds} | -2,5 | +0,8 |

Аналогично, по таблице 2 видно, что у p -канального транзистора при повышении температуры до +180°C уменьшаются коэффициенты вариации порогового напряжения и коэффициента модуляции длины канала, но увеличиваются коэффициенты вариации крутизны и выходной проводимости. При понижении температуры до -70°C наблюдаем повышение коэффициента вариации всех исследуемых параметров, кроме выходной проводимости.

Уменьшение коэффициента вариации говорит о том, что значения параметров имеют более низкую изменчивость, а их разброс уменьшается. Если коэффициент вариации возрастает, то разброс значений параметров увеличивается.

SPICE моделирование ВАХ транзисторов

Для определения параметров SPICE-модели был выбран метод *golden die*. С помощью средств обработки данных выводилась некоторая «средняя ВАХ», которая строилась по среднему значению точек ВАХ всех измеренных транзисторов. Далее выбирался транзистор, ВАХ которого была наиболее близка к средней характеристике. По этому транзистору и определялись параметры SPICE-модели семейства *BSIM4*. В SPICE модели

использовались встроенные параметры, с помощью которых подгонялись значения порогового напряжения, крутизны, коэффициента модуляции длины канала и выходной проводимости. Для учета воздействия температуры использовался блок параметров модели *BSIM4 Temperature Modeling Parameters*.

На рис.3-4 приведено сравнение измеренных и смоделированных ВАХ NFET-транзистора. Точность соответствия смоделированных характеристик с экспериментальными не превышает 15% во всем температурном диапазоне.

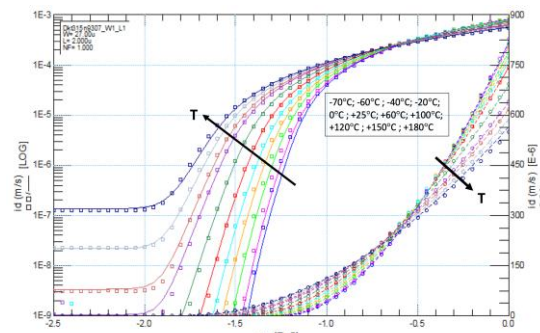


Рис.3. Сравнение измеренных (точки) и смоделированных (сплошная линия) сток-затворных ВАХ NFET-транзистора при $U_{cu}=4$ для разных температур от -70 до +180 °C в логарифмическом и линейном масштабах

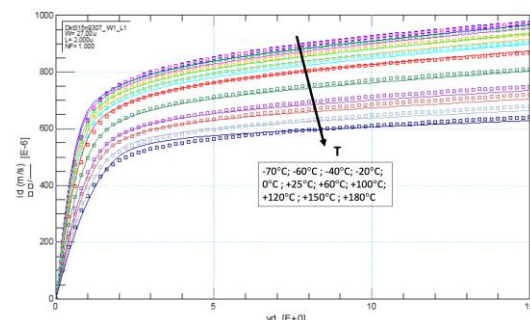


Рис.4. Сравнение измеренных (точки) и смоделированных (сплошная линия) выходных ВАХ NFET-транзистора при $U_{зи}=0$ для разных температур от -70 до +180 °C

Заключение

Проведён статистический анализ основных параметров JFET-транзисторов с учетом влияния температуры. Определены параметры SPICE-модели, учитывающие данное воздействие, и проведена их верификация.

Список литературы

1. Денисенко В. В. Компактные модели МОП-транзисторов для SPICE в микро- и нанoeлектронике. – М.: ФИЗМАТЛИТ, 2010. – с. 246.
2. P. A. 't Hart, M. Babaie, E. Charbon Subthreshold Mismatch in Nanometer CMOS at Cryogenic Temperatures // ESSDERC. 2019. P. 1-6.
3. P. A. 't Hart, J. P. G. van Dijk, M. Babaie Characterization and Model Validation of Mismatch in Na-

nometer CMOS at Cryogenic Temperatures // ESSDERC. 2018. P. 246-249.

4. M. Babaie, A. Vladimirescu, F. Sebastiano Characterization and Modeling of Mismatch in Cryo-CMOS // IEEE Journal of the Electron Devices Society. 2020. Vol. 8. P. 263-273.

РАЗРАБОТКА ПОРТАТИВНОГО ЭКСТРУДЕРА ДЛЯ ПРОИЗВОДСТВА ФИЛАМЕНТА

*Бадалян Р.К., Ключев Н.С.,
Мотайленко И.А., Цуркан А.А.*
*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Объектом исследования и разработки является портативный шнековый экструдер для производства пластикового филамента для 3D-принтеров. Готовый продукт можно эксплуатировать в учебных заведениях, на малых и крупных предприятиях и в бытовых условиях, чтобы экономить до 20% процентов материала, перерабатывая образовавшиеся в ходе работы 3D-принтера отходы. Также у пользователей есть возможность самостоятельно производить материал из пластиковых гранул, таким образом снизив затраты на покупку пластика более, чем в четыре раза. В результате работы разработан прототип экструдера: определен метод экструзии пластика, разработан модуль управления экструдером, представляющий из себя печатную плату, разработано программное обеспечение (далее ПО).

Введение

На данный момент на российском рынке не представлены бытовые экструдеры с полноценными печатными платами для управления, все аналоги состоят из отдельных функциональных частей, никак не связанных друг с другом. Предлагаемое решение (модуль управления экструдером) позволяет наладить связь между составными частями, чтобы добиться полного автоматизированного процесса производства пластиковой нити. Также в данной работе решается задача отсутствия температурных алгоритмов для управления кольцевыми нагревателями.

Целью работы является разработка и исследование устройства для бытового производства филамента для 3D-принтеров. Разработка состоит из нескольких этапов:

1. Изучение отрасли аддитивных технологий;
2. Выбор модуля управления;
3. Анализ перерабатываемого материала;
4. Схемотехническое моделирование;
5. Разработка ПО;
6. Разработка конструкции устройства;
7. Маркетинговые исследования;
8. Разработка документации.

Разрабатываемое устройство является вспомогательным модулем для 3D-принтера и в сфере промышленности и технологий относится к аддитивным технологиям. На 2023 год объем рынка аддитивных технологий оценивался в \$23,2 млрд., и ожидается, что совокупный годовой темп роста (CAGR - Compound annual growth rate) составит 22,4% на период с 2023 по 2026 год. Эксперты прогнозируют, что к 2026 году объем рынка аддитивных технологий составит \$43,5 млрд., причем 20% этой суммы отводится на разработку вспомогательных модулей и сервисов [1]. Однако и в России сфера аддитивных технологий стремительно растет с каждым годом. Согласно прогнозам правительства, объем российского рынка аддитивных технологий на 2022 г. составил 3,776 млрд. рублей, на 2025 прогнозируется рост до 6 млрд. рублей, в 2030 – 11,883 млрд. рублей. Совокупный среднегодовой темп роста составляет 13,2%. Доля отечественного рынка в данной сфере составляет всего 2%, Россия находится на 11 месте в мире [2].

Функционал экструдера

У большинства экструдеров есть существенный недостаток, а именно: отсутствие автоматизации, так как все функциональные части (нагреватель, датчик температуры, двигатель и т. д.) работают независимо друг от друга и никак не синхронизированы. Поэтому в данной работе было принято решение собрать полноценный модуль управления со всеми структурными частями экструдера. Устройство должно обладать следующими функциональными возможностями:

- Нагревать пластик до необходимой температуры плавления;
- Выдавливать расплавленный материал через сопло с круглым сечением;
- Автоматически наматывать полученную нить на катушку.

Уравнение диммера тока

Важным этапом исследования является вывод уравнения для преобразования мощности диммера. Изначально кольцевой нагреватель спроектирован так, что его нельзя нагреть до требуемой температуры. Интенсивность нагрева и мощность можно регулировать при помощи диммера тока, для этого предназначена библиотека “RBDdimmer.h”, однако там нет прямой зависимости между мощностью нагревателя и температурой, так как его можно подключить к любому устройству, ток которого можно изменять. Поэтому появляется необходимость самостоятельно провести исследование и вывести формулу, которая будет преобразовывать значение мощности диммера в значение температуры. Для этого проводилось исследование в полностью готовом корпусе экструдера: сначала интегрировался нагреватель в систему, затем к нему подключался NTC-термистор (Negative Temperature Coefficient), после чего активировалась работа нагревателя при мощности от 10–90% с шагом 10%. После этого выписывалось значение температуры для каждого значения мощности (выписывалось два значения: температура внутри камеры экструдера

и температура, считываемая термистором), полученные результаты представлены в таблице 1.

Таблица 1. Зависимость температуры от мощности диммера

| Мощность диммера | Температура на термисторе | Температура внутри камеры экструдера |
|------------------|---------------------------|--------------------------------------|
| 10 | 30 | 38 |
| 20 | 47 | 82 |
| 30 | 79 | 122 |
| 40 | 121 | 164 |
| 50 | 154 | 189 |
| 60 | 177 | 208 |
| 70 | 187 | 221 |
| 80 | 211 | 256 |

Затем строился график для двух значений и выводилось уравнение. Вдоль каждой зависимости выстраивалась полиномиальная линия тренда, так как максимально приближена к значениям (рис.1).

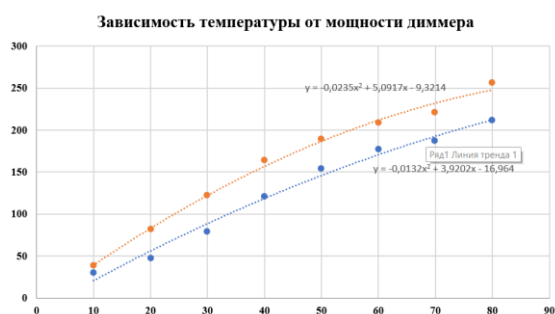


Рис.1. График зависимости температуры от мощности диммера

При работе с разрабатываемым устройством пользователю необходимо будет задавать именно температуру, от которой и будет зависеть установленная на нагревателе мощность. В этом случае в качестве аргумента будет выступать температура, а график принимает следующий вид (рис.2):

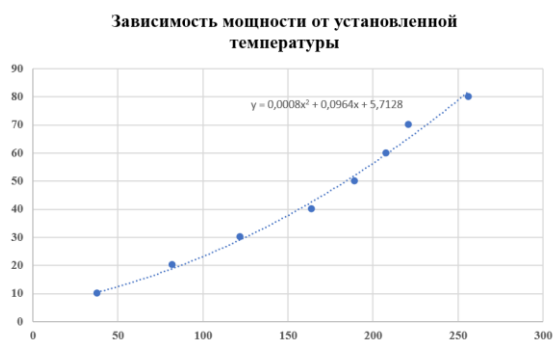


Рис.2. Зависимость мощности нагревателя от установленной пользователем температуры

В итоге получаем уравнение, которое отображает значение мощности от температуры, заданной пользователем (1)

$$P_{wr} = 0,0008 T_0^2 + 0,0964 T_0 + 5,7128 \quad (1)$$

Где T_0 – температура, установленная пользователем; P_{wr} – мощность диммера

Однако данное уравнение нельзя использовать в коде прошивки по нескольким причинам. Во-первых, на графике видно, что линия не проходит через точки, а в некоторых ситуациях погрешность составляет 5°C , что недопустимо в данной работе. Во-вторых, в качестве параметра мощности диммера можно устанавливать только натуральные числа до «100» и «0», а в полученном уравнении результат не всегда равен целому числу, поэтому придется округлять, что также недопустимо, так как погрешность получится еще больше. В-третьих, эксперимент проводился лишь при восьми значениях аргумента из ста возможных. Такое количество аргументов гарантировано приведет к неточностям и несоответствию уравнения прямой.

Для исправления этой ситуации было принято решение изменить алгоритм ПО и заменить функционал следующим образом: пользователь перед активацией системы получал рекомендации к работе с устройством и несколько значений температуры, которым соответствовало конкретное значение мощности, данные представлены в таблице 2.

Таблица 2. Значения мощности диммера и соответствующие значения температур

| Мощность диммера (%) | Температура нагревателя ($^\circ\text{C}$) |
|----------------------|--|
| 40 | 122 |
| 50 | 164 |
| 60 | 189 |
| 70 | 208 |
| 80 | 221 |
| 90 | 256 |

Разработка печатной платы и корпуса

Разработка печатной платы (далее ПП) – это достаточно долгий процесс, включающий в себя множество этапов, каждый из которых регламентирован в различных государственных и международных стандартах. Правильная разработка ПП (печатная плата) является одним из самых важных процессов в проектировании любой системы, так как именно ПП отвечает за связь и управление всеми составными частями устройства, и при повреждении ПП, вся система неизбежно выйдет из строя. Разрабатываемый прибор, для которого проектируется печатная плата, относится к стационарным приборам и согласно ГОСТу Р МЭК 335-1-94 [3] относится к приборам класса I. А сама конструкция относится к конструкциям класса III, так как плата питается безопасным сверхнизким напряжением.

Первый этап разработки ПП включает разработку принципиальной схемы, то есть расстановку всех электронных компонентов и назначение связей между ними. В первую очередь необходимо начать с разводки контроллера управления ATmega2560, затем необходимо поочередно добавить все составные части (резонаторы, преобразователи напряжений и т. д.). На рис.3 представлена принципиальная схема ПП.

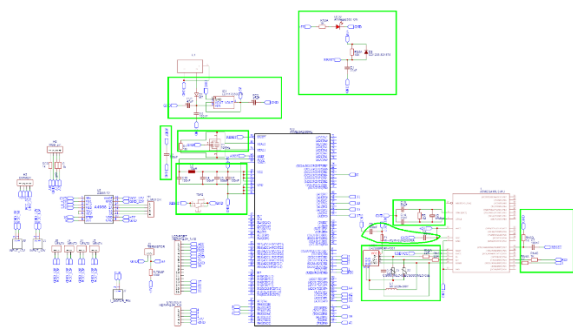


Рис.3. Принципиальная схема печатной платы

После разработки принципиальной схемы необходимо развести топологию печатной платы. В первую очередь необходимо выбрать элементную базу всей платы, рассчитав электрическую нагрузку для каждого компонента, а затем определить коэффициент электрической нагрузки (2):

$$K_n = \frac{F_{\text{раб}}}{F_{\text{ном}}}, \quad (2)$$

где $F_{\text{раб}}$ – электрическая нагрузка элемента в рабочем режиме; $F_{\text{ном}}$ – номинальная или предельная по ТУ (технические условия) электрическая нагрузка элемента, выполняющего в конструкции функцию схемного элемента.

Этот коэффициент в той или иной мере влияет на корректное функционирование и безотказность всех используемых элементов. В качестве параметра нагрузки выбирают характеристику, которая больше влияет на безотказность компонента. В данной работе больше всего подвергается нагрузке драйверы шаговых двигателей, так как они потребляют ток 1,5 А, параметром нагрузки для них является рассеиваемая мощность, а коэффициент равняется 0,31. Оптимальные значения нагрузки лежат в пределах от 0,2 до 0,6. [4]. Для определения электрической нагрузки необходимо провести симуляцию принципиальной схемы в среде Proteus (рис.4) и найти такие компоненты, характеристики которых допускают их использование в ПП.

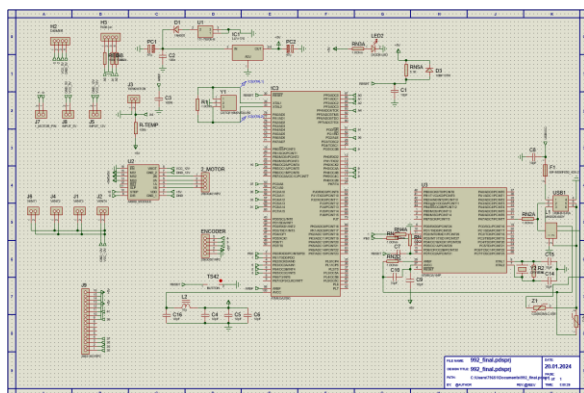


Рис.4. Симуляция в Proteus

Коэффициента электрической нагрузки компонентов недостаточно для их установки на печатную

плату, так как при работе любого электронного компонента на нем выделяется электрическая мощность, которая переходит в тепловую. У каждого компонента есть предельное значение температуры, при которой они могут исправно работать в случае, если температура используемого компонента превысит это значение, его придется заменить, добавить охлаждение или пересмотреть расстановку и конфигурацию платы. На рис.5–7 представлены верхний и нижний слой платы, а также ее 3D-модель.

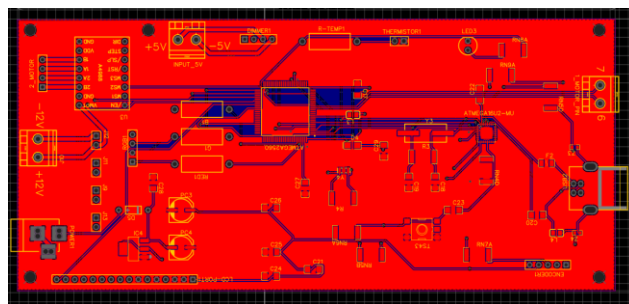


Рис.5. Верхний слой платы

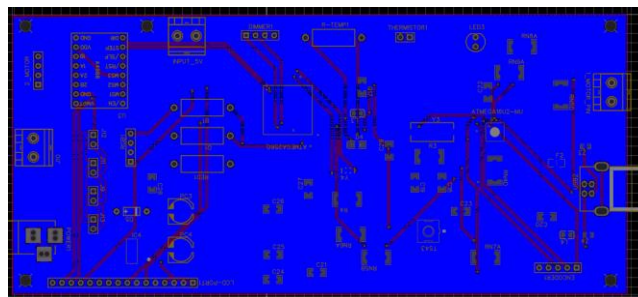


Рис.6. Нижний слой платы

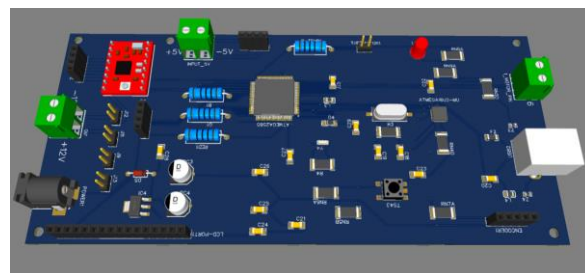


Рис.7. D-модель платы

Для проведения теплового моделирования необходимо экспортировать 3D-модель ПП в среду SolidWorks, так как там доступно тепловое моделирование, затем указываем материал для каждого компонента и указываем параметры моделирования: вид теплообмена – конвекция, а коэффициент конвективной теплоотдачи – 200 Вт/м². На рис.8 представлена изготовленная печатная плата.

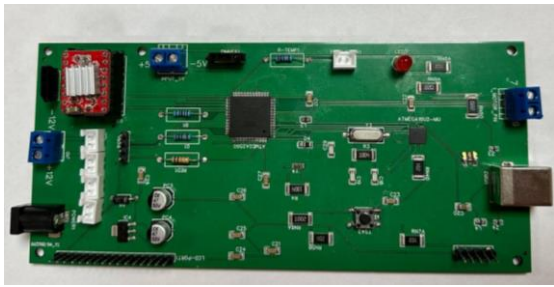


Рис.8. Тепловое моделирование платы

После изготовления печатной платы ее необходимо проверить на выделение тепла, для этого подключим плату к питанию и воспользуемся тепловизором (рис.9–10).

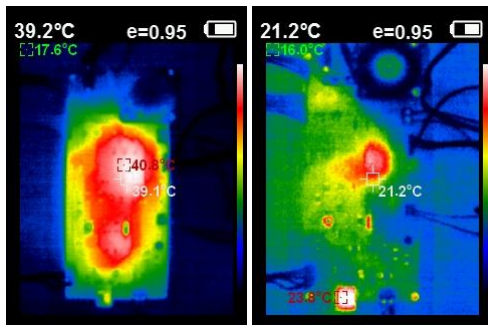


Рис.9 – 10. Рисунок с тепловизора

На рисунке видно, что больше всего греется контроллер ATmega2560 и модуль ATmega16U до 40°C, хоть производителем и была заявлена максимальная температура 85°C, было принято решение добавить кулер для охлаждения (рис.10) в таком варианте температура упала до 21°C.

Корпус устройства должен иметь достаточные габариты, чтобы внутри поместился экструдер, шаговый двигатель с редуктором, катушка для принтера диаметром 200 мм, а также изготовленная ПП, изолированная от экструдера, чтобы избежать перегрева. Корпус состоит из основания, корпуса экструдера и держателя катушки, такая конфигурация очень удобна для обслуживания и чистки принтера, а также позволяет легко получить доступ ко всем функциональным частям устройства. Корпус спроектирован таким образом, чтобы соответствовать классу безопасности IP30 согласно ГОСТу 14254–96 [5], так как защищена от попадания твердых предметов диаметром 2,5 мм и незащищена от воды (рис. 11).

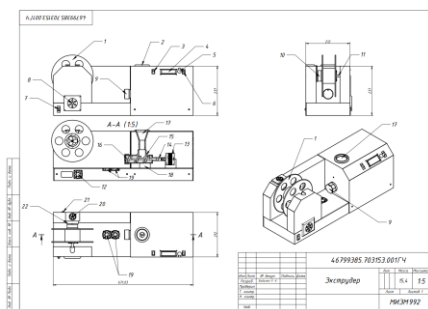


Рис.11. Габаритный чертеж экструдера

Заключение

В ходе выполнения работ было спроектировано устройство для производства филамента. Готовое решение состоит из двух частей: экструдера и катушки для наматывания и хранения пластиковой нити. Все функциональные части управляются с печатной платы, которая расположена в защитном корпусе экструдера. Также были описаны алгоритмы для преобразования мощности в температуру для диммера тока и выбран алгоритм для NTC-термистора. Для готового технического решения был получен патент на полезную модель [6].

Список литературы

1. D Printing Market Size, Share and Trends Analysis by Component Type, Vertical Type, Region and Segment Forecast to 2026 – URL: <https://www.globaldata.com/store/report/3d-printing-market-analysis/>
2. Распоряжение Правительства Российской Федерации от 14 июля 2021 г. № 1913-р «СТРАТЕГИЯ развития аддитивных технологий в Российской Федерации на период до 2030 года, Приложение №1» // Утверждена распоряжением Правительства Российской Федерации (Москва)
3. ГОСТ Р МЭК 335-1-94. Безопасность бытовых и аналогичных электрических приборов. Общие требования и методы испытаний. Safety of household and similar electrical appliances. General requirements and test methods. Дата введения 1995-01-01 / Разработан и ВНЕСЕН Техническим комитетом по стандартизации бытовых электроприборов (ТК 19). Утвержден и введен в действие Постановлением Госстандарта России от 2 марта 1994 г. N 43
4. Боровиков С. М., Троян Ф. Д., Цырельчук И. Н. Расчет показателей надежности радиоэлектронных средств // Министерство образования Республики Беларусь Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники» - Минск. – 2010. – С. 12–13.
5. ГОСТ 14254–96 (МЭК 529–89). Межгосударственный стандарт степени защиты, обеспечиваемые оболочками (код IP) Degrees of protection provided by enclosures (IP Code). Дата введения 1997-01-01 / Разработан Техническим комитетом по стандартизации ТК 341 "Внешние воздействия". Внесен Госстандартом России. Принят Межгосударственным советом по стандартизации, метрологии и сертификации (протокол N 9 от 12 апреля 1996 г.)
6. Патент RU 219926 U1 Российская Федерация МПК В29В 17/00, Устройство для переработки отходов пластика после 3D-печати / Бадалян Р. К., Мотайленко И. А., Сафронов М.А., Эфендиев С.В; Патентообладатель: федеральное государственное автономное образовательное учреждение высшего образования "Национальный исследовательский университет "Высшая школа экономики" (RU). – N 2023108600, заявл. 05.04.2023; опубл. 14.08.2023; Бюл. №23.

кость рефрактометра. Это изображение также видит и исследователь (врач), который по характеру полученного изображения может выявить дальнозоркость (или же гиперметропию) или близорукость (миопию) – болезни, возникающие в связи с неправильной фокусировкой световых лучей на сетчатке глаза. При миопии центр фокусировки лучей находится как бы «раньше» сетчатки, а при дальнозоркости наоборот – «за» сетчаткой (см. рис.2).

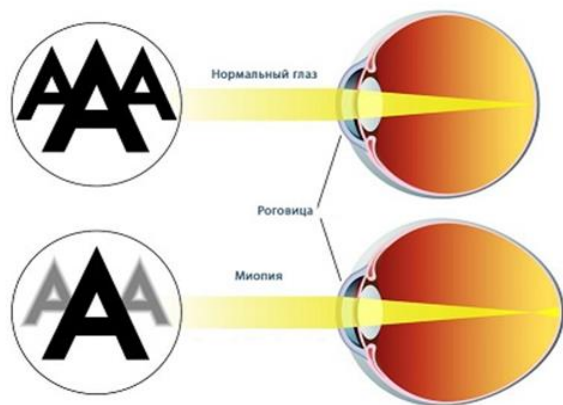


Рис.2. Правильная фокусировка лучей на сетчатке глаза (эмметропия) и неправильная при близорукости

При помощи упомянутого выше прибора рефрактометра также можно выявить *астигматизм* – дефект зрения, который возникает вследствие неравномерной кривизны полусферы роговицы глаза, из-за чего свет не фокусируется в одной точке, а рассеивается, что и приводит к нечеткому зрению (см. рис.3).



Рис.3. Разница между фокусировкой световых лучей на сетчатке глаза при нормальном зрении и астигматизме

Далее перейдем к осмотру *глазного дна*. Всего выделяют около 20 заболеваний, которые можно обнаружить при обследовании глазного дна: от катаракты до сахарного диабета, склероза и артрита. Всё из-за того, что глазное дно – это единственная часть тела человека, где можно напрямую увидеть нервные волокна и сосуды. Осмотр глазного дна проводится при помощи офтальмоскопа, который, отправляя и

принимая отраженный от сетчатки световой луч, способен пятнадцатикратно увеличить изображение тканей глаза. По полученному изображению врач способен выявить симптомы различных патологий.

Измерение *внутриглазного давления* проводится чаще всего бесконтактно (такой процесс называется пневмотонометрией). В ходе исследования прибор создает поток воздуха, направленный в глаз пациента, после чего считывает изменение формы роговицы как ответную реакцию на процесс. По изменению роговицы прибор получает итоговое давление глаза.

Оценить *сферичность роговицы глаза* помогает лазер, при помощи которого создается своеобразная топографическая «карта», где различным способом обозначены участки с разной высотой.

Когда врач, проведя соответствующий осмотр, подозревает патологические процессы в глазу, проводят исследование *критической частоты слияния световых мельканий (КЧСМ)*. Суть этого метода диагностики заключается в определении критической частоты световых вспышек, при которой сетчатка остается в состоянии воспринимать их по отдельности, а не как единый световой поток. Различные заболевания, в том числе атрофия нерва, дистрофия сетчатки, глаукома, склероз, вызывают снижение частоты слияния мельканий в разной степени. Что важно, значения КЧСМ практически не зависят от остроты зрения, величины зрачка и рефракции глаза, что позволяет проводить исследование при соответствующих дефектах зрения.

Цветовосприятие человека традиционно проверяют при помощи таблиц Рабкина, предлагая назвать изображенные геометрические фигуры или цифры, образованные кружочками разных цветов.

Поскольку обыкновенный смартфон достаточно ограничен в своем функционале, и не может, к примеру, создать лазер и в дальнейшем принимать отраженный от сетчатки глаза луч, в работе при помощи специальных устройств и смартфона с VR-очками оцениваются следующие показатели зрения человека: состояние глазного дна, цветовосприятие, острота зрения и КЧСМ.

Описание структуры системы

Система состоит из трех ключевых частей:

Первая часть представляет собой бинокулярный офтальмоскоп, фиксирующийся на смартфон. Электроника и источники света обеспечивают съемку изображений глазного дна. Мобильное приложение, использующее нейронную сеть, обученную на датасете заболеваний, классифицирует состояние глазного дна.

Вторая часть основана на технологии виртуальной реальности (VR). Разработано мобильное приложение, голосовой ассистент и алгоритмы проверки цветовосприятия и остроты зрения. Пользователь может проводить тестирование, взаимодействуя с голосовым ассистентом через VR-очки, интегрированные с мобильным телефоном.

Третья часть включает прибор для исследования критической частоты слияния световых мельканий (КЧСМ). Этот инструмент, связанный с мобильным

приложением через Bluetooth, позволяет пользователю самостоятельно оценить работу сетчатки глаза, а также выявить дальтонизм.

На рис.4 представлена схема взаимодействия составных частей системы.

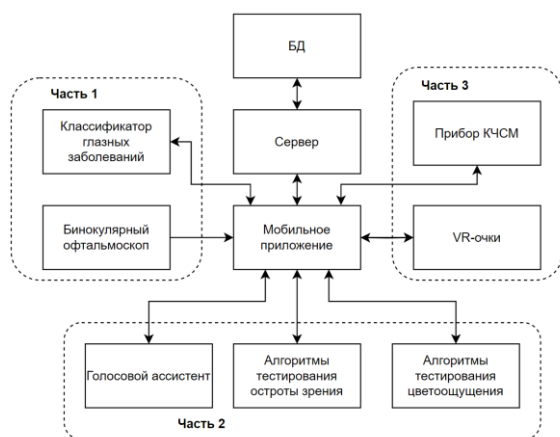


Рис.4. Схема взаимодействия составных частей системы

Описание голосового помощника и его роль в системе

Голосовой помощник интегрируется с целью обеспечить удобное взаимодействие с помощью голосовых команд на мобильном устройстве, предназначенном для оценки ключевых показателей зрения. Это позволит пользователям минимизировать необходимость взаимодействия с интерфейсом и сконцентрироваться на процессе проверки зрения. Голосовое управление позволит пользователям легко и эффективно взаимодействовать с системой, не отвлекаясь на традиционные методы ввода данных [5]. Основными характеристиками голосового помощника будут встроенная поддержка русского языка и возможность распознавания как отдельных букв, так и непрерывной речи. Голосовой помощник будет работать с различными шаблонами команд, что позволит пользователям легко и быстро получать необходимую информацию или выполнять определенные действия. Важным аспектом функционала голосового помощника будет его готовность к возможному расширению функционала и обслуживания на различных площадках.

Функционал мобильного приложения

Мобильное приложение в данной работе выступает как средство, при помощи которого в VR-очках можно проверить свою остроту зрения при помощи таблицы Сивцева и цветоощущение по таблицам Рабкина. Помимо этого подразумевается, что в мобильном приложении пациент как основной пользователь обладает возможностью зарегистрировать свой собственный аккаунт, где будет храниться история диагностических исследований, которая в дальнейшем может быть отражена у врача. Но только в том случае, если клиника, к которой отнесен врач, зарегистрировала в приложении супераккаунт

для себя и своих сотрудников. Для помощи пользователю (пациенту) в интерпретации результатов его диагностики, приложение по завершении сеанса исследования будет формировать список советов и рекомендаций пациенту, поскольку оно, в отличие от врача-специалиста, не может и не должно самостоятельно ставить диагнозы.

Архитектура сети

На рис.5 изображена архитектура сети.



Рис.5. Архитектура сети

Заключение

Совершенствование технических решений для проверки показателей зрения человека является актуальным направлением согласно новым вызовам современности.

В рамках исследования изучена предметная область, разработан комплекс оценки показателей зрения человека: программное обеспечение (мобильное приложение) для контроля остроты зрения, цветоощущения и КЧСМ, программный модуль анализа состояния глазного дна с точностью до 93%.

В дальнейшем планируется создание макета VR-очков с дополнительной оптической системой для контроля состояния глазного дна, интеграция голосового помощника в мобильное приложение, а также внедрение продукта на рынок устройств для улучшения процессов диагностики, мониторинга и лечения пациентов с офтальмологическими заболеваниями.

Список литературы

- Worldwide prevalence and risk factors for myopia // Wiley Online Library URL: <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/j.1475-1313.2011.00884.x> (дата обращения: 13.02.2024).
- Classification of Eye Diseases in Fundus Images // IEEE URL: <https://ieeexplore.ieee.org/ielx7/6287639/9312710/09474508.pdf?tp=&arnumber=9474508&number=9312710&ref=> (дата обращения: 13.02.2024).
- Результаты исследования цветового зрения разными способами у детей с амблиопией // Российская Детская Офтальмология URL: <https://rdo.su/index.php/rdo/article/view/164> (дата обращения: 13.02.2024).
- КЧСМ в определении зрительной работоспособности слабовидящих школьников // Институт проблем передачи информации им. А.А. Харкевича (ИППИ) РАН URL: <http://www.jip.ru/2002/GOL1.pdf> (дата обращения: 13.02.2024).
- Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors // ScienceDirect URL: <https://www.sciencedirect.com/science/article/pii/S0969698920312911> (дата обращения: 13.02.2024).

РАЗРАБОТКА ИНСТРУМЕНТОВ ВЕРИФИКАЦИИ САПР ЦИФРОВЫХ СБИС

Грицун Т.Ю.², Кашинцев И.К.², Портнягин Е.Р.¹,
Ржевская М.М.², Чижевмар М.М.¹

Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент электронной инженерии,

²департамент прикладной математики

МИЭМ НИУ ВШЭ

Аннотация

В статье рассматривается задача автоматизированной проверки корректности компонентов САПР логического синтеза цифровой аппаратуры. Представлены описания разработанных средств и результаты их тестирования.

Введение

В контексте развития отечественной микроэлектроники важной задачей является обеспечение технологической независимости путем разработки собственных систем автоматизации проектирования (САПР) сверхбольших интегральных схем (СБИС). Важным этапом процесса проектирования цифровой микроэлектронной аппаратуры является логический синтез — построение логической схемы в заданном технологическом базисе по исходной RTL-модели (Register Transfer Level), представленной на языке описания аппаратуры (например, на языке Verilog [1]).

На этапе логического синтеза проводятся следующие преобразования:

1) трансляция RTL-модели в промежуточное представление логической схемы;

2) оптимизация промежуточного представления по заданным критериям (площадь, общая задержка, энергопотребление);

3) отображение схемы на элементы технологической библиотеки (технологическое отображение).

Проверка корректности (верификация) этих преобразований критически важна для функционирования средств логического синтеза. Важной задачей здесь является проверка логической эквивалентности (Logic Equivalence Checking, LEC) представлений синтезируемой схемы (в том числе, до и после проведения оптимизаций).

Проверяемые представления считаются логически эквивалентными, если 1) они имеют одинаковые интерфейсы (совпадают как количества первичных входов и выходов соответственно, так и их разрядности); 2) при подаче на входы представлений любых идентичных значений на выходы также передаются одинаковые значения. Так как перебор всех возможных входных значений для СБИС является чрезвычайно затратным процессом, используются специализированные методы LEC.

Методы проверки логической эквивалентности

Существует ряд подходов к проверке эквивалентности логических схем: 1) метод Мищенко [2], 2) метод Кюльмана [3], 3) метод, основанный на

двоичных диаграммах решений (Binary Decision Diagram, BDD) [4]. Общей подзадачей для данных методов является построение митер-схемы или митры (miter). Митра — это композиция двух логических схем. Митра имеет тот же входной интерфейс, что и сравниваемые схемы; значения со входов митры передаются на одноименные входы схем. Значения с одноименных выходов схем попарно сравниваются таким образом, что если хотя бы на одном наборе входных значений сравниваемые схемы вернут разные значения хотя бы на одной паре выходов, то митра вернет значение «истина» (true), свидетельствующее о том, что сравниваемые схемы не эквивалентны.

В контексте необходимости разработки отечественных средств логического синтеза актуальной задачей является сравнительная оценка эффективности методов LEC. Для решения этой задачи были разработаны следующие программные реализации методов LEC.

- Компонент, основанный на симуляции митры (RND). На основании исходных схем строилась митра. Затем проводилась её симуляция на конечном наборе случайных значений, либо на всех возможных значениях (применимо для небольших схем).

- Компонент, основанный на комбинационном варианте алгоритма Мищенко (Functionally Reduced AIG, FRAIG). На основании исходных схем строилась митра и преобразовывалась в представление в базисе AIG (И, НЕ). В результате симуляции множество внутренних вентилях схемы разбивалось на классы эквивалентности. Для пары вентилях одного класса строились максимальные конусы — подсхемы, состоящие из тех вентилях исходной схемы, из которых заданный вентиль достижим. Затем SAT-решателем MiniSAT [5] проверялась эквивалентность построенных конусов. В случае подтверждения их эквивалентности первый вентиль в паре удалялся из исходной митры, затем удалялись вершины, ставшие «висячими» (не связанные с первичными выходами схемы). При этом второй вентиль заменял собой первый в тех частях схемы, в которых первый фигурировал в качестве входа для других вентилях. После перебора всех классов вентилях по сокращенной (reduced) митре строилась логическая формула и проверялась с помощью SAT-решателя.

- Компонент BDD, основанный на конвертировании логической схемы в BDD-представление. На основании исходных схем строилась митра, которая затем передавалась компоненту, реализующему конвертацию. Если в процессе конвертации митра преобразовывалась в константу 0, то это свидетельствовало об эквивалентности исходных схем.

Методология тестирования

Тестирование проводилось с целью анализа эффективности разработанных прототипных реализаций и проверки согласованности полученных результатов. Помимо вышеописанных программных реализаций, в тестировании также был задействован компонент проверки эквивалентности, основанный на SAT-решателе MiniSAT.

Реализации сравниваемых методов ЛЕС были разработаны на языке C++ и интегрированы в прототип среды логического синтеза, разрабатываемой в ИСП РАН.

В качестве тестового набора использовались Verilog-описания из набора тестов для парсера logina [6] (10 примеров), а также реализации базовых арифметических операций (сложение, вычитание, умножение) разных разрядностей (6, 16) на внутреннем языке представления комбинационных схем (5 примеров). Характеристики схем указаны в таблице 1. В ходе тестирования по RTL-моделям строились промежуточные представления логических схем. Для проведения экспериментального сравнения компонентов проверки эквивалентности использовались как позитивные тесты, так и негативные. Для позитивных тестов использовался подход с использованием отображения на один из логических базисов: AIG (И, НЕ), XAG (исключающее ИЛИ, И, НЕ), MIG (функция голосования, НЕ, константа 0), XMG (исключающее ИЛИ, функция голосования, НЕ, константа 0); отображенная схема сравнивалась с исходной. В негативных тестах использовался специальный программный компонент, называемый мутатором логических схем. Данный компонент принимал на вход представление логической схемы и модифицировал её заведомо неэквивалентным образом (строит схему-мутанта). Схемы-мутанты сравнивались с исходными схемами с помощью компонентов проверки эквивалентности; предполагалось, что на данных тестах проверка даст негативный результат.

Для сравнения с существующими инструментами проверки логической эквивалентности использовался OpenABC (ABC) [7]. Проводилось отображение, затем исходная схема сравнивалась с отображенной. Использовались те же схемы из набора тестов для парсера logina (10 примеров). Результаты представлены в таблице 4.

Тестирование проводилось на компьютере с 8 Гб ОЗУ и 4-ядерным процессором с частотой 4 ГГц. В ячейках таблиц 2 и 3 указано время работы компонентов в миллисекундах. Символ “-” означает, что в схеме более 64 входов, и проверку эквивалентности провести не удалось в силу ограничений прототипных реализаций. Символ “х” означает, что результат не был получен за отведенные 10 минут. Данное ограничение является аппроксимацией ожидаемого предельного допустимого времени работы компонентов с учетом размеров рассматриваемых схем.

Таблица 1. Характеристики схем

| Схемы | Входы | Внутр. вентили | Язык |
|-------|-------|----------------|---------|
| c17 | 5 | 32 | Verilog |
| c432 | 36 | 621 | Verilog |
| c499 | 41 | 1 083 | Verilog |
| c880 | 60 | 1 082 | Verilog |
| c1355 | 41 | 1 595 | Verilog |
| c1908 | 33 | 1 397 | Verilog |
| adder | 256 | 5 102 | Verilog |

| Схемы | Входы | Внутр. вентили | Язык |
|-------------|-------|----------------|-----------|
| arbiter | 256 | 47 211 | Verilog |
| bar | 135 | 9 132 | Verilog |
| sub | 34 | 428 | вн. язык |
| add | 34 | 412 | вн. язык |
| add6 | 14 | 204 | вн. язык |
| sub6 | 14 | 158 | вн. язык |
| mul6 | 14 | 515 | вн. язык |
| Balanc eAND | 5 | 20 | синтетич. |
| AndOr | 4 | 16 | синтетич. |
| square | 64 | 70 912 | Verilog |

Таблица 2. Результаты позитивного тестирования

| Схемы | Метод | | | |
|---------|---------|------|---------|-------|
| | SAT | RND | FRAIG | BDD |
| c17 | 1 | 33 | 121 | 4 |
| c432 | 288 | х | 3 382 | 3 212 |
| c499 | 630 | х | 73 878 | 233 |
| c880 | 886 | х | 34 159 | 749 |
| c1355 | 761 | х | 211 881 | 3 679 |
| c1908 | 833 | х | 221 160 | 408 |
| adder | 2 127 | - | - | х |
| arbiter | 148 121 | - | - | 1 712 |
| bar | 150 383 | - | - | х |
| sub | 33 | х | 188 | 10 |
| add | 31 | х | 345 | 13 |
| add6 | 18 | 603 | 120 | 3 |
| sub6 | 14 | 637 | 97 | 3 |
| mul6 | 136 | 2383 | 1 548 | 7 |

Таблица 3. Результаты негативного тестирования

| Схемы | Метод | | | |
|-------------|-------|-----|--------|-------|
| | SAT | RND | FRAIG | BDD |
| adder | 64 | - | - | х |
| arbiter | 1 316 | - | - | 1 793 |
| bar | 33 | - | - | 3 024 |
| c17 | 1 | 1 | 5 | 4 |
| c499 | 3 | 6 | 21 174 | 1 071 |
| c1908 | 192 | 11 | 67 061 | 343 |
| Balance AND | 1 | 1 | 8 | 3 |
| AndOr | 1 | 1 | 11 | 3 |
| square | 5 913 | - | - | х |

Таблица 4. Результаты позитивного тестирования инструмента ABC

| Схемы | Время работы ABC, ms |
|-------|----------------------|
| c17 | 10 |
| c432 | 10 |
| c499 | 10 |
| c880 | 50 |
| c1355 | 20 |

| Схемы | Время работы ABC, ms |
|---------|----------------------|
| c1908 | 20 |
| adder | 10 |
| arbiter | 10 |
| bar | 210 |
| square | 3 120 |

По результатам тестирования на позитивных тестах реализация метода SAT оказалась наиболее эффективной с точки зрения быстродействия: для 11 примеров время работы не превышало одной секунды. Второй по быстродействию была реализация метода BDD: для 9 примеров время работы не превышало 1 секунды. Стоит отметить схемы (adder, bar), на которых результат за отведенное время получен не был, поскольку размер BDD-представления увеличивался слишком быстро. Метод RND оказался наиболее быстрым при тестировании на схемах-мутантах: для всех примеров, на которых удалось провести проверку эквивалентности, время работы не превышает 0.1 секунды. Инструмент ABC оказался быстрее представленных методов на всех схемах, кроме самой маленькой. Стоит отметить, что реализация метода проверки эквивалентности, основанного на BDD, является единственной не зависящей от количества входов схемы, что позволяет выполнять проверку схем с большим (> 64) количеством входов наиболее эффективно. С точки зрения сравнения с ABC

Заключение

Проверка корректности преобразований, в частности отображения на логический базис, являющегося разновидностью технологического отображения, критически важна для функционирования средств логического синтеза. В рамках работы сравнивались прототипы компонентов проверки эквивалентности, использующие следующие подходы: SAT, RND, FRAIG, BDD. На позитивных тестах наиболее эффективным с точки зрения быстродействия оказался прототип SAT, на негативных — прототип RND.

Основными направлениями дальнейшей работы являются отладка и оптимизация разработанных прототипов. Также планируется расширить множество сравниваемых компонентов проверки эквивалентности за счет реализации метода Кюльмана. Кроме того, планируется расширить тестовый набор схемами из открытых бенчмарков ISCAS'85 [8] и EPFL [9].

Список литературы

1. The Institute of Electrical and Electronics Engineers IEEE Std 1364™-2005 (Revision of IEEE Std 1364-2001) IEEE Standard for Verilog® Hardware Description Language. - Текст : непосредственный // April 2006, ISBN 0-7381-4850-4.
2. Mishchenko, A. Improvements to combinational equivalence checking / A. Mishchenko, S. Chatterjee, R. Brayton, N. Eem. - Текст : непосредственный // ICCAD, 2006.

3. Kuehlmann, A. Equivalence checking using cuts and heaps / A. Kuehlmann, F. Krohm. - Текст : непосредственный // *Proceedings of Design Automation Conference (DAC)*, pp. 263-268, 1997.

4. Lee, C. Y. Representation of Switching Circuits by Binary-Decision Programs / C. Y. Lee. - Текст : непосредственный // *Bell System Technical Journal*, Vol. 38, July 1959, pp. 985-999.

5. Репозиторий MiniSAT : [сайт]. – 2024. - URL: <https://github.com/niklasso/minisat> (Дата обращения: 29.10.2024). – Текст : электронный.

6. Репозиторий lorina : [сайт]. – 2024. - URL: <https://github.com/hriener/lorina> (Дата обращения: 29.10.2024). – Текст : электронный.

7. Инструмент ABC : [сайт]. – 2024. - URL: <https://people.eecs.berkeley.edu/~alanmi/abc/> (Дата обращения: 22.06.2024). – Текст : электронный.

8. Hansen, M. C. Unveiling the ISCAS-85 benchmarks: A case study in reverse engineering /

M. C. Hansen, H. Yalcin, J. P. Hayes. - Текст : непосредственный // *IEEE Design & Test of Computers*, vol. 16, no. 3, pp. 72–80, July–Sept. 1999.

9. Репозиторий EPFL : [сайт]. – 2024. - URL: <https://github.com/lsils/benchmarks> (Дата обращения: 29.10.2024). – Текст : электронный.

АЛГОРИТМ ОБРАТНОЙ СВЯЗИ РУЛЕВОГО УПРАВЛЕНИЯ ДЛЯ VR-СИМУЛЯТОРА ЯХТЫ

Стрельцов Г.А., Юлдашов А.С.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Статья посвящена разработке и анализу алгоритма обратной связи для VR-симулятора яхты. Основное внимание уделено принципам управления шаговыми моторами и их интеграции с технологией Field Oriented Control (FOC) для создания алгоритма обратной связи для VR-симулятора яхты. Алгоритм обратной связи обеспечивает реалистичное взаимодействие пользователя с виртуальной средой.

Введение

Основная цель исследования заключается в создании алгоритма, который обеспечивает реалистичные ощущения от управления яхтой, тем самым улучшая качество обучения и погружение пользователя в процесс. Яхтинг становится все более популярным среди любителей и профессионалов, что требует высококвалифицированной подготовки и безопасных условий обучения. Одной из ключевых проблем является необходимость практического опыта, сопряженного с рисками и затратами. VR-симуляторы предлагают инновационное решение, создавая реалистичные и безопасные условия для тренировки. Алгоритмы обратной связи обеспечивают реалистичное усилие от штурвала, значительно

повышая эффективность обучения. В первом разделе обсуждается логика обратной связи на штурвале, во втором разделе - существующие подходы к созданию алгоритма и их недостатки, в третьем разделе - базовая основа моделирования обратной связи, в четвертом - выбор мотора для обратной связи, в пятом - выбор метода управления шаговым мотором, в шестом - практическая реализация алгоритма обратной связи.

Обратная связь в реальном рулевом управлении яхты

Для того, чтобы понять, за счет чего при рулении на штурвале оказывается силовое воздействие [1], препятствующее повороту штурвала, необходимо рассмотреть принцип работы рулевого управления парусного судна. Для корректировки поворота яхты применяется рулевое перо. Оно представляет из себя плоскую или профилированную пластину, расположенную за кормой. Вращение штурвала вызывает поворот рулевого пера на соответствующий угол. В прямом положении штурвала перо располагается параллельно направлению движения, но при изменении угла меняется площадь воздействия воды на перо, что в свою очередь вызывает смену направления движения яхты. При этом, чем больше площадь пера, на которую воздействует вода, тем большая сила передается на штурвал. За счет вышеописанной физики процесса величина обратной связи на штурвале при разном направлении движения будет разной. Так, при движении судна прямо сопротивление повороту штурвала будет небольшим и будет нарастать с увеличением угла поворота штурвала. При движении яхты назад вода не оказывает сопротивление, а наоборот способствует повороту штурвала. При этом сила, помогающая повернуть штурвал при движении назад, больше, чем сила, оказывающая сопротивление повороту штурвала при движении вперед.

Существующие подходы для создания обратной связи рулевого управления

Проведенный обзор показал, что основными подходами решения задачи, подобной рассматриваемой нами, являются использование открытой MMOS прошивки на типовой DIY-системе рулевой силовой обратной связи из руля, энкодера и мотора, описанной в [2]; также использование прошивок от компаний на готовых гоночных симуляторах [3]; также использование установки из дрели [4] и энкодера и вшитой программе для Arduino. В алгоритмах прошивок и программ запрограммировано возвращающееся рулевое усилие, однако основная проблема всех этих подходов заключается в том, что программы не воспроизводят уникальные особенности управления яхтой. В частности, при движении яхты задним ходом реальное управление характеризуется наличием корректирующего усилия в сторону поворота руля. Этот фактор играет ключевую роль в реалистичном воссоздании ощущений от управления

яхтой и должен быть интегрирован в алгоритм обратной связи. В своей работе мы постарались найти решение данной проблемы.

Моделирование обратной связи

При разработке алгоритма обратной связи была использована упрощенная модель, учитывающая лишь основные аспекты гидродинамики, так как главной задачей работы была техническая реализация базовой модели обратной связи. Внесение более точной физической модели является потенциалом для развития работы.

Для того, чтобы вычислить необходимый момент силы, который должен создаваться на валу шагового мотора для обеспечения правдоподобной обратной связи, необходимо учесть, что штурвал соединяется с рулевым пером не напрямую, а через привод управления и рулевую передачу [5]. Введем понятие передаточного коэффициента K_{Π} , который является отношением силы, действующей на штурвал ($F_{\text{удерж}}$ к силе, действующей на перо руля ($F_{\text{сопр}}$). Данный коэффициент можно вычислить, как отношение максимального угла поворота пера в одну сторону (40°) к максимальному углу поворота штурвала (355°). Таким образом, $K_{\Pi} \approx 0,112$. Формула момента силы, который должен быть приложен к штурвалу для обеспечения обратной связи, будет иметь следующий вид (1):

$$M = K_{\Pi} \cdot F_{\text{сопр}} \cdot r, \quad (1)$$

где M - момент силы, который должен создаваться на валу шагового двигателя для обеспечения правдоподобной обратной связи на штурвале, r - радиус штурвала (в данном случае $0,5\text{м}$), K_{Π} - передаточный коэффициент, $F_{\text{сопр}}$ - величина силы, прикладываемой к виртуальному перу руля.

Как было показано выше, сила $F_{\text{сопр}}$ - это сила, с которой вода будет давить на рулевое перо. Для вычисления необходимого момента силы, который должен создавать двигатель, и подбора комплектующих необходимо было определить, какую силу нужно прикладывать к штурвалу, чтобы обратная связь на нем казалась правдоподобной. Для этого был проведен эксперимент (см. рис.1), в котором с одной стороны штурвала опытный яхтсмен оказывал нужное воздействие, с другой стороны были подвешены эталонные грузики. Из этого был получен максимальный удерживающий момент, исходя из которого была вычислена оптимальная сила обратной связи и подобран необходимый для ее реализации мотор.

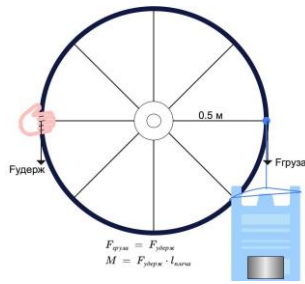


Рис.1. Эксперимент с грузами

В результате эксперимента было измерено, что максимальная сила $F_{удерж}$ для правдоподобной обратной связи на штурвале тренажера равняется $\sim 40\text{Н}$ при движении яхты вперед. В реальном мире эта сила состоит из силы лобового сопротивления воды, а также сторонних сил, связанных с течением воды, таких как сила, придаваемая течением или волнами. Для упрощения модели на начальном этапе проектирования тренажера было принято решение пренебречь сторонними силами, так как в условиях небольших скоростей и спокойного состояния воды, малых кренов, представляющих собой основной случай для симуляции на тренажере, другие составляющие малы. Сила лобового сопротивления [6] вычисляется по формуле (2):

$$F_{сопр} = C_x \cdot \frac{\rho v^2}{2} \cdot S, \quad (2)$$

где C_x - коэффициент лобового сопротивления, ρ - плотность среды, v - скорость перемещения объекта в среде, S - площадь движущегося объекта.

Исходя из приведенной выше формулы, коэффициент лобового сопротивления можно вычислить по формуле (3), в которую подставляются значения силы воздействия среды на площадь миделева сечения объекта в этой среде. Данная сила определяется с помощью эксперимента. В данном случае используется значение силы, полученной в вышеописанном эксперименте с грузами. Для того, чтобы вычислить значение коэффициента лобового сопротивления для рассматриваемой задачи, необходимо подставить в формулу (3) значения силы $F_{сопр}$, исходя из вычисленной экспериментально $F_{удерж}$, ρ - плотности воды, значения максимальной скорости яхты v_{max} , площади пера руля $S \cdot \sin(40^\circ)$ (максимальное сопротивление будет, когда перо отклонено на максимальный угол поворота 40°):

$$C_x = \frac{2F_{сопр}}{\rho v_{max}^2 \cdot S \cdot \sin(40^\circ)}, \quad (3)$$

где $S = 0,2\text{м}^2$.

При движении вперед $v_{max} \approx 3 \frac{\text{м}}{\text{сек}}$. Получаем значение $C_x \approx 0,62$. При движении назад экспериментально измеренная сила $F_{удерж} \approx 60\text{Н}$

($F_{сопр} \approx 536\text{Н}$, $v_{max} \approx 1 \frac{\text{м}}{\text{сек}}$). Соответственно, для движения назад $C_x \approx 0,92$. Так как для данной задачи рассматриваются небольшие скорости, то будем считать, что значения C_x будут одинаковы при разных скоростях судна и углах поворота штурвала. Итоговая формула для расчета момента будет иметь вид (4):

$$M = C_x \cdot \frac{\rho v^2}{2} \cdot S \cdot r \cdot \sin\alpha, \quad (4)$$

где α - угол поворота рулевого пера. При положительных значениях момента перо будет отклонено вправо, при отрицательных влево.

Зависимость момента от угла поворота пера и скорости движения яхты (для примера рассматривается движение судна вперед и $C_x = 0,04$) можно увидеть на следующем графике (см. рис.2):

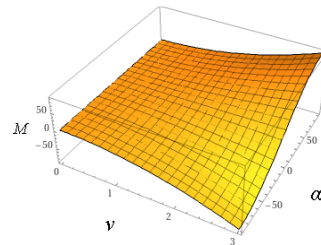


Рис.2. Зависимость момента от скорости и угла поворота пера

Выбор мотора для создания обратной связи

Для разработки алгоритма обратной связи на штурвале яхты необходимо было тщательно выбрать подходящий тип мотора (см. рис.3), способного обеспечить высокую точность позиционирования, простое управление, надежность и экономичность.

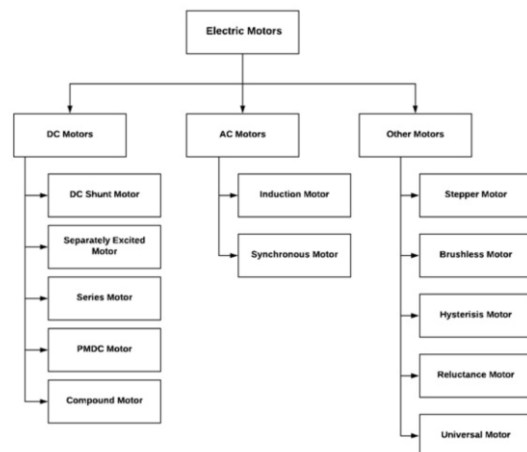


Рис.3. Типы электрических моторов

После анализа различных типов электрических моторов [7] было решено отказаться от мотора постоянного тока (DC) и мотора переменного тока (AC) из-за их ограниченной точности, сложности

управления. Подобная обратная связь реализуется в большинстве своем с помощью бесколлекторных двигателей постоянного тока (BLDC) моторов, так как они удобны, надежны и доступны по цене. Шаговый двигатель был выбран как представитель BLDC, имеющий выше создаваемый момент на малых скоростях, чем у обычных BLDC, и более доступные комплектующие.

Выбор метода управления шаговым мотором

При выборе мотора для системы обратной связи, важно учитывать несколько ключевых характеристик и требований, чтобы обеспечить максимальную реалистичность и эффективность системы. Во-первых, мотор должен обладать высокой точностью и чувствительностью, что позволит точно воспроизводить различные сценарии управления, и обеспечивать адекватную обратную связь. Вторым важным аспектом является способность мотора быстро реагировать на изменения, что критически важно для имитации динамически изменяющихся условий управления яхтой. Также мотор должен обладать достаточным моментом силы для воспроизведения реалистичных усилий, испытываемых при управлении яхтой. Это включает в себя способность имитировать как легкие, так и более сильные усилия на штурвале. Кроме того, важен широкий диапазон рабочих скоростей, позволяющий имитировать различные условия управления, от медленных и плавных до быстрых и резких маневров. С учетом этих требований, одним из наиболее подходящих методов управления мотором для системы обратной связи является использование метода Field-Oriented Control (FOC) [8]. FOC или векторное управление, представляет собой современный метод управления электродвигателями, центральным элементом которого является управление магнитным полем внутри двигателя. Позволяет раздельно контролировать магнитный поток и момент двигателя, обеспечивая тем самым высокую точность и эффективность управления.

Практическая реализация алгоритма обратной связи

Была использована open-source библиотека для управления шаговыми и BLDC двигателями «SimpleFOC» [9], которая позволяет реализовать векторное управление. С помощью данной библиотеки можно задавать целевой момент силы для двигателя с помощью имплементированного в библиотеку алгоритма FOC. Также для тестирования алгоритма был собран стенд, состоящий из контроллера ESP8266 (подключается к ПК, как HID устройство), драйвера шагового двигателя L298N, поворотного энкодера AMT102-v, который закрепляется на валу двигателя, шагового мотора Nema 42 и источника питания на 24В. Схема подключения компонентов изображена на рис.4.

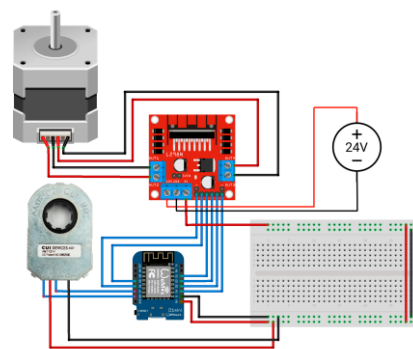


Рис.4. Подключение компонентов стенда

Компоненты для тестирования подбирались из следующих соображений: микроконтроллер ESP8266 имеет высокую процессорную частоту и поддерживается библиотекой SimpleFOC, драйвер L298N также поддерживается данной библиотекой и позволяет управлять каждой фазой мотора по отдельности, что необходимо для управления мотором через FOC. Энкодер AMT102-v необходим для реализации алгоритма FOC, а также позиционирования поворота штурвала. Данный энкодер обладает высокой точностью, что позволяет наиболее корректно вычислять необходимый вектор поля для FOC. Nema 42 обладает удерживающим моментом в 30Н·м, который является достаточным для создания правдоподобной обратной связи.

Библиотека SimpleFOC предоставляет набор методов для реализации алгоритма FOC. Для шагового двигателя данная библиотека реализует специальный класс StepperMotor, который предоставляет возможность управлять фазами мотора с помощью 4PWM драйвера. С использованием данной библиотеки нами была разработана программа, реализующая полный алгоритм обратной связи рулевого управления яхты, блок-схема которого представлена на рис.5.

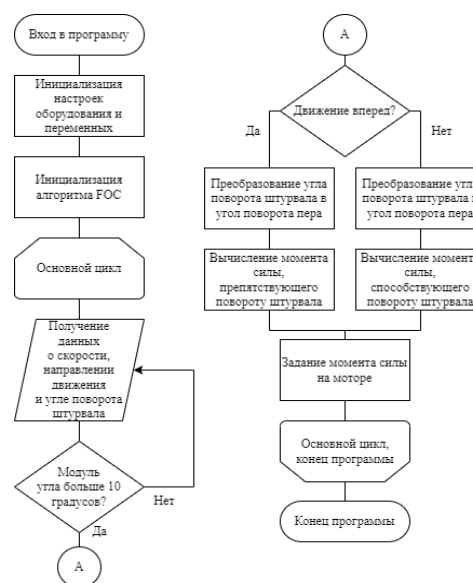


Рис.5. Алгоритм обратной связи рулевого управления

Данная библиотека использует синусоидальную модуляцию для управления ШИМ сигналами, подаваемыми на фазы электромотора. Также для шагового мотора имплементирован режим “управления напряжением”. Данный подход к управлению вращательным моментом позволяет устанавливать целевое напряжение U , а алгоритм FOC вычисляет необходимые фазовые напряжения U_α и U_β для обеспечения плавной работы. Данные напряжения затем распределяются по 4 фазам шагового мотора и устанавливаются с помощью 4 выходов ШИМ драйвера.

Для того, чтобы вычислить, какое напряжение необходимо передавать в метод, отвечающий за управление моментом, необходимо воспользоваться данной формулой, которая описывает зависимость момента силы от тока (5):

$$M = k_t \cdot I, \quad (5)$$

где M - целевой момент силы, k_t - постоянная момента силы мотора, I - ток, проходящий через обмотки мотора. С помощью данной формулы и закона Ома можно вычислить необходимое напряжение (6):

$$U = R \cdot \frac{M}{k_t}, \quad (6)$$

где R - сопротивление обмоток мотора. R и k_t можно найти в даташите к используемому мотору или измерить экспериментально.

Заключение

На созданном нами стенде для проверки для проверки алгоритма обратной связи, описанного в данной статье подхода, были проведены испытания с привлечением опытного яхтсмена, это тестирование показало высокий уровень реалистичности передаваемого на штурвал усилия обратной связи.

Интеграция шаговых моторов и алгоритма FOC в VR-симулятор яхты создала эффективную, точную систему управления для реалистичного виртуального взаимодействия. Это значительно повышает потенциал симулятора для обучения, обеспечивая глубокое погружение в опыт управления яхтой.

Список литературы

1. Багрянцев Б.И., Решетов П.И. "Учись морскому делу". Москва: ДОСААФ, 1986. С. 26.
2. Granite Devices. MMos [Электронный ресурс]. – URL: <https://granitedevices.com/wiki/MMos> (дата обращения: 10.02.2024).
3. Driver Easy. How to Set Up a Gaming Steering Wheel on Windows 7, 8 and 10 [Электронный ресурс]. – URL: <https://www.drivereasy.com/knowledge/set-up-gaming-steering-wheel-on-windows-7-8-and-10/> (дата обращения: 10.02.2024).
4. Winfield, Jason. DIY Force Feedback Steering Wheel Using a Cordless Drill Motor [Электронный ресурс]. – URL: [https://jasonwinfield.nz/diy-force-](https://jasonwinfield.nz/diy-force-feedback-steering-wheel-using-a-cordless-drill-motor)

[feedback-steering-wheel-using-a-cordless-drill-motor](https://jasonwinfield.nz/diy-force-feedback-steering-wheel-using-a-cordless-drill-motor) (дата обращения: 10.02.2024).

5. Кацман Ф.М., Дорогостайский Д.В., Коннов А.В., Коваленко Б.П. "Теория и устройство судов". Ленинград: Судостроение, 1991. С. 122.

6. Войткунский Я.И. "Сопротивление движению судов". Издание 2. Ленинград: Судостроение, 1988. С. 22.

7. ElProCus. Different Types of Electric Motors [Электронный ресурс]. – URL: <https://www.elprocus.com/different-types-of-electric-motors/> (дата обращения: 10.02.2024).

8. Performance Motion Devices. Field Oriented Control (FOC) - A Deep Dive [Электронный ресурс]. – URL: <https://www.pmdcorp.com/resources/type/articles/get/field-oriented-control-foc-a-deep-dive-article> (дата обращения: 10.02.2024).

9. Arduino SimpleFOCLibrary [Электронный ресурс]. – URL: https://docs.simplefoc.com/arduino_simplefoc_library_showcase (дата обращения: 10.02.2024).

МОДЕРНИЗАЦИЯ СТРУКТУРЫ ОПРОСНИКА ОЦЕНКИ КОЭФФИЦИЕНТА КАЧЕСТВА ПРОИЗВОДСТВА ЭЛЕКТРОННЫХ МОДУЛЕЙ НА ОСНОВЕ СПРАВОЧНИКА “HANDBOOK OF 217PLUS RELIABILITY PREDICTION MODELS” ДЛЯ ПРОГРАММНОГО ИНСТРУМЕНТА

*Бушуев Н.И., Козлова А.А.,
Мкртчян Г.А., Насыров Д.Д.*

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе приводится анализ существующей системы оценки коэффициента качества производства электронных модулей, которая основана на американском справочнике. Однако данный способ содержит определенные недостатки: для корректного результата необходимо ответить на огромное количество вопросов, при этом они не систематизированы и часто близки по смыслу между собой. По этой причине нами предложен усовершенствованный опросник по оценке коэффициента качества производства электронных модулей, который будет реализован в виде программного инструмента.

Введение

Современные технологические решения и инженерные разработки на производствах сталкиваются с высокими требованиями к надежности. К сожалению, в нынешних реалиях необходимые условия обеспечения надежности в большинстве производственных организациях не выполняются. Для улучшения надежности нужно выполнять конкретные процедуры, уровень выполнения которых можно оценить с помощью коэффициента качества произ-

водства (K_a) [1, 2, 3].

На данный момент оценку коэффициента качества производства можно провести с помощью справочника “Handbook of 217plus reliability prediction model” [4], который является одним из широко используемых инструментов для данной задачи, однако зачастую его прохождение вызывает затруднения из-за большого количества несистематизированных и однотипных вопросов, а также отсутствует программное решение для данного опросника.

Исходя из выявленных недостатков, нами была выдвинута следующая цель – усовершенствовать прохождение опроса для его реализации в виде программного решения. Для этого мы решили структурировать вопросы справочника, а также применить идею автоматических ответов для них. В первом разделе описана текущая система оценки качества производства предприятия по данному справочнику. Во втором разделе представлена модификация опросника.

Существующий способ оценки коэффициента качества производства

Один из способов оценки коэффициента качества производства предприятия K_a строится на справочнике “Handbook of 217plus reliability prediction model”. В нем есть данные и формулы, которые позволяют рассчитать программную, прогнозируемую (данные для их оценки даны в справочнике), а также базовую интенсивность отказов. Для расчета последнего вида интенсивности необходимо проанализировать и оценить 7 категорий (причин отказа), а именно:

1. P_D - конструктивные отказы;
2. P_M - производственные отказы;
3. P_P - отказы комплектующих элементов;
4. P_S - несовершенство системы управления;
5. P_N - несовершенство методов контроля;
6. P_I - эксплуатационные отказы;
7. P_W - деградационные отказы;
8. P_G - варвар несовершенство управления повышением надежности.

Каждый из них (кроме P_G , для его расчета смотреть формулу 3) оценивается по формуле:

$$P_i = \alpha_i (-\ln(R_i))^{\frac{1}{\beta_i}} \quad (1),$$

где α и β - постоянные, которые даны в справочнике, а R_i - рейтинг процесса для i -ой категории (рассчитывается по формуле 2).

$$R_i = \frac{\sum_{j=1}^{n_i} G_{ij} W_{ij}}{\sum_{j=1}^{n_i} W_{ij}} \quad (2),$$

где G_{ij} - ответ для j -ого вопроса, i -ой категории отказа; W_{ij} - вес вопроса j -ого вопроса, i -ой категории отказа; n_i - количество вопросов для конкретной категории.

Значение параметров G_{ij} и W_{ij} берутся из соответствующих таблиц справочника. Пример формы продемонстрирован в таблице 1.

Таблица 1. Значение коэффициентов и параметров при расчете параметра R метрики P_G

| № п/п | Вопрос | G_{ij} | W_{ij} |
|-------|---|-----------------------------|----------|
| 1 | Существует ли эффективная система передачи сообщений об отказах и анализа внесения исправлений (FRACAS) для развернутой системы? | Yes = 1 No = 0 | 8 |
| 2 | Каков процент отказов в процессе эксплуатации, для которых определена первопричина? | $G = \text{percentage}/100$ | 8 |
| 3 | Проводится ли анализ, чтобы определить, повторяется ли сбой? | Yes = 1 No = 0 | 6 |
| 4 | Определены ли потенциальные корректирующие действия, связанные с проектированием, производством или управлением системой? | Yes = 1 No = 0 | 6 |
| 5 | Проводятся ли консультации с первоначальными разработчиками или производственным персоналом по поводу возможных внесенных исправлений? | Yes = 1 No = 0 | 4 |
| 6 | Существует ли инфраструктура поддержки на местах, способная повлиять на необходимые изменения? | Yes = 1 No = 0 | 10 |
| 7 | Адекватно ли протестированы системы, чтобы убедиться, что изменения были сделаны правильно и не привели к другим дефектам или повреждениям? | Yes = 1 No = 0 | 5 |

Из анализа опросника, который был создан для эксперта, следует, что порядок вопросов для разных коэффициентов не систематизирован, и вопросы не подразделяются по каким-либо критериям.

Помимо этого, следует отметить, что данная система имеет большое количество вопросов (их более 450), что делает систему затруднительной для прохождения, к тому же многие вопросы однотипны и включают ответы на другие.

Предлагаемый метод оценки коэффициента качества производства

Модифицированная система оценки K_a заключается в распределении вопросов в соответствии их принадлежности к определенным уровням электронного модуля [5] (уровень ЭМ) и категориям предметной области, которые уникальны для каждого вида отказов (категория предметной области), а так-

же во введении автоматических ответов на некоторые вопросы (предполагаемый ответ), которые будут использоваться программным инструментом.

Таблица 2. Распределение вопросов, относящихся к метрике Π_G , согласно предлагаемому методу

| Вопрос | Уровень ЭМ | Категория предметной области | Предполагаемый ответ |
|--|------------|------------------------------|----------------------|
| Существует ли эффективная система передачи сообщений об отказах и анализа внесения исправлений (FRACAS) для развернутой системы? | 3, система | FRACAS | - |
| Каков процент отказов в процессе эксплуатации, для которых определена первопричина? | все | Анализ | - |
| Проводится ли анализ, чтобы определить, повторяется ли сбой? | все | Риски/опыт | - |
| Определены ли потенциальные корректирующие действия, связанные с проектированием, производством или управлением системой? | 3, система | Техподдержка | G |

Распределение элементов опроса подобным образом позволяет отследить взаимосвязи между ними, определить систему внутри каждой части опроса. Это в свою очередь позволит применить в программном решении технику автоматических ответов, которые будут зависеть от того, какой ответ дал респондент на нижестоящем в иерархии вопросе. Пример разбиения продемонстрирован в таблице 2 с вопросами, относящихся к оценке несовершенства управления повышением надежности, где G значит, что ответ предполагается соответствующим нижестоящему, а \bar{G} – обратным. Также в таком случае появляется возможность введения комбинированных вопросов, требующих комплексные ответы - список, последовательность и др. Стоит отметить, что распределение элементов системы оценки в соответствии с принадлежностью к тем или иным уровням электронного модуля, позволяет при предварительном выборе сотрудником производства, проходящего опрос, сферы деятельности производства убрать из опроса ряд вопросов нерелевантных для предприятия. В таком случае предполагается, что программой автоматически будет выставлен ответ $G = 0$.

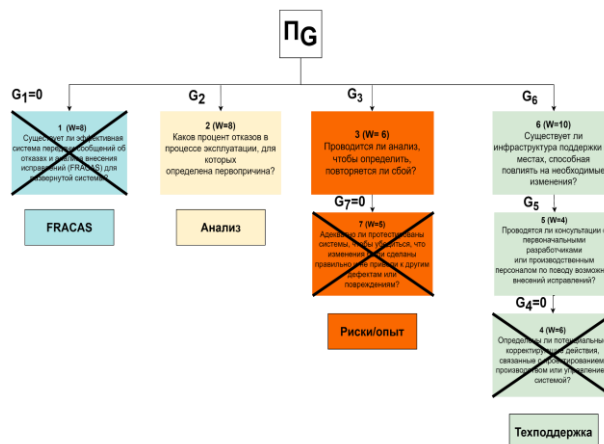


Рис. 1. Дерево части опроса, относящаяся к несовершенству управления повышением надежности, адаптированная под предложенный пример

Визуализация работы модифицированной системы оценки представлена на рис. 1 в виде дерева, где в качестве примера взята часть опроса, относящаяся к анализу несовершенства управления повышением надежности для производства, занимающегося 2 уровнем ЭМ. Данный случай наглядно показывает, что для оценки показателя достаточно ответить на 4 из 7 вопросов, а расчет метрики в соответствии с таблицей 1 сводится к следующей формуле:

$$\Pi_G = \frac{1,12(t+2)^{-\alpha}}{2^{-\alpha}} \quad (3),$$

$$\text{где } \alpha = R = \frac{\sum_{j=1}^7 G_j W_j}{\sum_{j=1}^7 W_j} = \frac{8 \cdot G_2 + 6 \cdot G_3 + 4 \cdot G_5 + 10 \cdot G_6}{47}$$

Данные нововведения позволяют улучшить процесс прохождения опроса в виде программного инструмента, давая возможность ответить только на часть вопросов, представленных в справочнике, что позволяет сделать опросник более доступным и понятным для респондентов.

Заключение

В ходе работы была рассмотрена существующая система оценки качества предприятия, основанная на справочнике “Handbook of 217plus reliability prediction model”. Система оценивания включает в себя ряд вопросов, разбитых на категории, а также соответствующие им весовые коэффициенты. Однако существующий опросник, построенный на вопросах для оценки, имеет следующие недостатки: вопросы не классифицированы для разных коэффициентов, некоторые из них включают в себя ответы на другие, а также нет программного решения для данного опроса.

В связи с существующими недостатками опросника была предложена модифицированная система оценки коэффициента качества производства, которая основывается на определении принадлежности вопроса к соответствующему уровню электронного модуля и категории предметной области. Кроме того,

программный инструмент, который будет создан на основе него, позволит получать ответы на ряд вопросов автоматически, что значительно облегчит проведение опроса.

Благодарность

Публикация подготовлена в ходе проведения исследования (Проект № 24-00-024 «Развитие методов прогнозирования показателей надежности электронных модулей») в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2024 г.

Список литературы

1. Артюхова, М. Оценка показателей надежности электронных средств с учетом многофакторного коэффициента качества производства / М. Артюхова, В. Жаднов, С. Полесский // Компоненты и технологии. – 2014. – № 4. – С. 204-207.

2. Королев, П. С. Исследование влияния коэффициента качества производства аппаратуры на оценку показателей надежности радиолокационного оборудования / П. С. Королев // Научные чтения по авиации, посвященные памяти НЕ Жуковского. – 2019. – С. 90-94.

3. Королев, П. С. Оценка „коэффициента качества производства “для модели интенсивности отказов радиотехнических приборов непиотируемых автоматических космических аппаратов / П. С. Королев, В. В. Жаднов // Известия высших учебных заведений. Приборостроение. – 2020. – Т. 63. – № 3. – С. 264-277.

4. Denson, W. Handbook of 217plus reliability prediction models / W. Denson. – RIAC, 2006.

5. Информационная технология многофакторного обеспечения надежности сложных электронных систем / Н. К. Юрков, А. В. Затылкин, С. Н. Полесский [и др.] // Надежность и качество сложных систем. – 2013. – № 4. – С. 75-79.

АНАЛИЗ МЕТОДОВ ОЦЕНКИ НАДЕЖНОСТИ СИСТЕМ ПРОМЫШЛЕННОГО ИНТЕРНЕТА ВЕЩЕЙ

Цветков В.Э., Ландер Л.Б.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Данное исследование посвящено разработке метода оценки надежности системы промышленного интернета вещей (IIoT), который позволит усовершенствовать процессы прогнозирования и контроля надежности устройств, предъявления требований к ним на этапах проектирования и производства. В статье проведен обзор используемых способов по расчету показателей надежности систем IIoT, про-

анализированы их недостатки. Более того, разработан собственный метод оценки надежности, позволяющий учесть как влияние аппаратной, так и программной и сетевой составляющих системы. Данное исследование проводится в рамках выполнения выпускной квалификационной работы по многофакторному прогнозированию показателей надежности автоматизированной системы управления (АСУ) IIoT.

Введение

За последнее десятилетие, по причине улучшения интернет-технологий и развития интеллектуальных персональных вычислительных устройств, применение интернета вещей (IoT), а также его приложений и поддерживающих аппаратных платформ стало актуальной темой как в научных, так и в практических кругах (рис.1) [1].

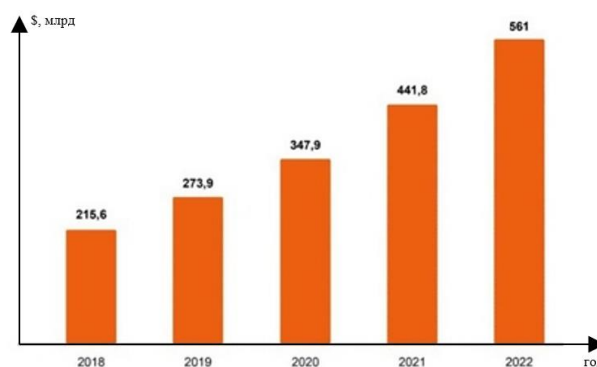


Рис.1. Рост рынка IIoT (данные Statistics MRC)

Системы IoT можно разделить на два класса [2]. К первому классу относится потребительский IoT, который включает в себя различные умные системы, используемые в бытовых условиях и позволяющие автоматизировать множество процессов в доме, оптимизировать энергопотребление и сделать жизнь человека более комфортной. Второй класс, включающий в себя приложения от мониторинга до автономного управления в промышленных секторах, называется IIoT.

В отличие от первого класса приложений, возникновение сбоев или отказов в эксплуатации любых промышленных системах могут привести к остановке всей производственной линии, повреждению инфраструктуры предприятия или даже к гибели работников.

Более того, в современных радиоэлектронных устройствах (РЭУ) существует проблема несоответствия результатов расчетной оценки показателей надежности на этапе проектирования с реальной статистикой отказов при эксплуатации на этапе производства. Причина данного несоответствия заключается в том, что существующие методы оценки надежности учитывают лишь аппаратную часть систем.

Все эти причины приводят к актуализации задачи обеспечения высокого уровня надежности в системах IIoT.

Цель работы данного исследования заключается в повышении точности оценки показателей надежности систем IoT. В ходе исследования была изучена архитектура систем IoT, рассмотрены способы расчета надежности таких систем, проанализированы их недостатки, а также предложен разработанный подход к многофакторному прогнозированию их показателей надежности.

Данная работа состоит из двух разделов. В первой части исследования представлена архитектура систем IoT, а также сводный отчет с обзором и анализом существующих методов их расчета надежности. Во второй части предлагается новый метод оценки показателей надежности IoT устройств, учитывающий техническую, ПО и сетевую составляющие.

Текущие подходы к расчету надежности IoT

Одним из подходов к оценке показателей надежности IoT является метод, представленный в американском справочнике MIL-HDBK-217F [3]. Согласно нему, надежность таких систем учитывает только техническую составляющую и рассчитывается по (1):

$$R_{IoT} = R_{hardware} \quad (1)$$

где R_{IoT} – надежность системы IoT, $R_{hardware}$ – надежность аппаратного обеспечения.

В свою очередь надежность составной части аппаратного обеспечения вычисляется по (2):

$$\lambda_p = \lambda_b * \pi_Q * \pi_T * \pi_S * \pi_A * \pi_E * \pi_R * \pi_C \quad (2)$$

где λ_p – интенсивность отказов составной части, λ_b – базовая интенсивность отказов составной части, учитывающая влияние электрической и температурной нагрузок, π_i – коэффициенты, учитывающие различные факторы (окружающая среда, качество и т.д.), которые влияют на надежность составной части.

Однако, данный подход не является корректным в связи с тем, что IoT – многоуровневая система, которая помимо датчиков и контроллеров, встроенных в узлы и устройства промышленных подсистем, имеет в своем составе также программную и сетевую компоненты, надежность которых оказывает не меньшее влияние на корректное функционирование системы (рис.2) [4]. Более того, необходимо учитывать влияние таких факторов, как каналы передачи информации, диапазон рабочих частот, режимы модуляции, протоколы, программное обеспечение и другие составные элементы IoT. Это поможет как заказчику, так и разработчику на этапе проектирования подобной системы оценивать показатели надежности с помощью многофакторной оценки, а не только опираться на техническую надежность оборудования при установке длительности гарантийного периода разработанной системы.



Рис.2. Архитектура IoT

Более комплексный подход к расчету надежности таких систем представлен в [5]. Авторы предлагают оценивать надежность IoT в соответствии с (3):

$$R_{system} = R_{hard}(t) * R_{soft}(t) * R_{human}(t) \quad (3)$$

где R_{system} – надежность системы IoT в целом, $R_{hard}(t)$ – надежность аппаратной части, $R_{soft}(t)$ – надежность программной части, $R_{human}(t)$ – надежность, учитывающая влияние человеческого фактора (HRA).

Для расчета надежности аппаратной части используется выражение (2), приведенное ранее. В свою очередь, надежность программной части можно оценить с помощью американского справочника RIAC-HDBK-217Plus [6]. Интенсивность отказов программного обеспечения λ_{SW} рассчитывается по (4):

$$\lambda_{SW} = \left(\frac{F_{t_{i-1}} - F_{t_i}}{730} \right) * (DC * FL * FA * AS) * 10^6 \quad (4)$$

где F_{t_i} – количество отказов за время t_i , DC – цикл работы программного обеспечения, FL – задержка при отказе, FA – активация отказов, AS – доля критичных отказов с точки зрения пользователя.

Тем не менее данный метод также не является полным, так как не учитывает сетевую составляющую надежности и в то же время учитывает человеческий фактор, являющийся опциональным в системах IoT.

Уточненный метод оценки надежности системы IoT

Проведя анализ недостатков существующих способов расчета надежности IoT, был разработан собственный подход к оценке показателей надежности систем IoT. Структурная схема расчета надежности (CxPH) системы IoT представлена на рис.3.

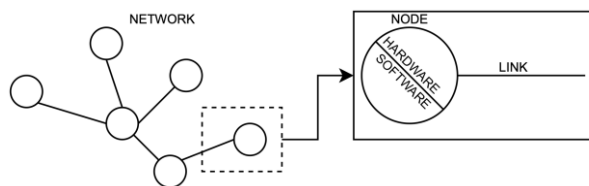


Рис.3. CxPH системы IoT

Согласно уточненному методу, надежность таких систем рассчитывается в соответствии с (5)-(7):

$$R_{IoT} = R_N \quad (5)$$

$$R_N = R_{node} * R_{link} \quad (6)$$

$$R_{node} = R_{HW} * R_{SW} \quad (7)$$

где R_N – надежность сети, R_{node} – надежность узла, R_{link} – надежность связи, R_{HW} – надежность аппаратного обеспечения, R_{SW} – надежность программного обеспечения.

Таким образом, надежность системы IoT является по сути надежностью ее сети. Надежность сети в свою очередь складывается из надежностей узлов и связей. Надежность узлов представляет собой надежность аппаратного и программного обеспечений. Аппаратная составляющая рассчитывается в соответствии со справочником «Надежность ЭРИ» 2006 года или MIL-HDBK-217F по формуле (2), представленной выше. Оценка программной составляющей производится по справочнику RIAC-HDBK-217Plus и приведенной ранее формуле (4). На расчет надежности связи влияет используемый тип связи (проводная или беспроводная). При проводном соединении для обеспечения высокого уровня надежности необходимо на этапе проектирования произвести расчет сечения кабеля. Соответственно, при этом виде связи нужно учесть такие факторы, как материал, сечение и длина проводника, существенно влияющие на величину потерь в линии. Если же в системе IoT задействована беспроводная связь, то для оценки ее надежности необходимо рассчитать вероятность битовой ошибки (BER) согласно [7]. Более того, при оценке надежности связи в разработанном методе учитываются такие параметры, как рабочий диапазон частот, в котором функционирует система, используемый тип модуляции, количество каналов связи, топология сети, а также вид протокола передачи данных.

Заключение

В ходе данного исследования были рассмотрены существующие подходы к оценке надежности IoT. Проведенный анализ показал, что данные методы недостаточно эффективны в связи с тем, что в них учитываются не все составляющие многоуровневой системы. Данный факт ведет к невозможности достоверного прогнозирования показателей надежности устройств IoT на этапах проектирования и разработки.

Для решения вышеописанной проблемы в данной работе был предложен собственный метод оценки надежности системы IoT, включающий как аппаратную, так и программную и сетевую компоненты. Данный подход позволит автоматизировать и усовершенствовать процессы формирования требований к проектированию и производству систем IoT и повысить их качество.

Благодарность

Публикация подготовлена в ходе проведения исследования (Проект № 24-00-024 «Развитие методов прогнозирования показателей надежности электронных модулей») в рамках Программы «Научный фонд

Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2024 г.

Список литературы

1. Промышленный интернет вещей. Практическое применение. - Текст: электронный // Smartgopro: сайт. – 2024. – URL: https://smartgopro.com/novosti2/IoT_tech/ (дата обращения: 13.02.2024)
2. Dong-Seong Kim, Hoa Tran-Dang, “Reliability Evaluation Model of Industrial Internet of Things Systems,” 1st International Conference on Engineering (ICONE 2020), vol. 1, No. 1, pp. 22-24, 2020.
3. Military Handbook. Reliability prediction of electronic equipment. MIL-HDBK-217F, 1991.
4. Industrial Internet of Things – IoT Промышленный интернет вещей. - Текст: электронный // TADVISER: сайт. - 2024. - URL: <https://www.tadviser.ru/index.php/IoT>
5. Slavko J. Pokorni. Reliability and availability of the internet of things // Vojnotehnički glasnik / military technical courier. 2019, Vol. 67, Issue 3, pp. 588-600.
6. RIAC-HDBK-217Plus. Handbook of 217PlusTM Reliability Prediction Models. USA: RIAC, 2006. 170 p.
7. Recommendation ITU-R F.1605. Error performance and availability estimation for synchronous digital hierarchy terrestrial fixed wireless systems. Switzerland: ITU, 2009. 13 p.

ИНСТРУМЕНТЫ ТРАНСЛЯЦИИ HDL-ОПИСАНИЙ ДЛЯ ИСПОЛЬЗОВАНИЯ В САПР

Мешочков Д.М.¹, Ушаков А.Н.²

Национальный исследовательский университет

«Высшая школа экономики»,

¹департамент электронной инженерии,

²департамент прикладной математики

МИЭМ НИУ ВШЭ

Аннотация

Работа направлена на создание открытых инструментов трансляции описаний функциональных устройств на языке Verilog [1] и технологических библиотек в формате Liberty [2]. Разрабатываемые подсистемы планируется использовать в САПР Utopia EDA, разрабатываемой ИСП РАН.

Введение

Для эффективного проектирования современной цифровой микроэлектроники широкое применение нашли САПР. Основным процессом САПР цифровых устройств является логический синтез, выполняющий несколько ключевых задач: построение внутреннего представления схемы, его оптимизации в соответствии с заданными критериями и отображение схемы на элементы технологической библиотеки.

На вход инструментам логического синтеза подается RTL-модель, которая создается на языке описания аппаратуры (HDL – hardware description

language), таком как Verilog. Затем RTL-модель транслируется во внутреннее представление САПР (Net), на основе которого осуществляется дальнейший анализ и трансформация модели.

Одним из этапов логического синтеза является технологическое отображение – процесс преобразования модели таким образом, чтобы она состояла из ограниченного набора элементов, задаваемых технологической библиотекой. В технологической библиотеке описываются базовые ячейки производства, их физические свойства: временные задержки, энергопотребление, занимаемая площадь и т.д. и их функциональность. Одним из распространенных форматов описания технологических библиотек является Liberty. Таким образом, для технологического отображения также требуется соответствующий транслятор.

Помимо этого, в рамках САПР Utopia EDA рассматривается возможность использования сторонних инструментов в качестве компонентов САПР. Одним из таких инструментов является проект CIRCT [3]. Наиболее простым способом интеграции CIRCT является взаимодействие посредством передачи моделей в одном из поддерживаемых проектом форматов, например, FIRRTL [4]. Для использования уже имеющихся RTL-моделей в формате Verilog в CIRCT необходимо иметь возможность трансляции из Verilog в FIRRTL.

Трансляция описаний технологических библиотек в формате Liberty

Транслятор реализован в виде программной библиотеки на языке C++. Основными компонентами транслятора являются:

- 1) лексический анализатор;
- 2) синтаксический анализатор;
- 3) анализатор ограничений.

Входом транслятора является файл описания технологической библиотеки в формате Liberty. Выходом транслятора является набор структур данных, описывающих содержимое технологической библиотеки: ячейки, их интерфейсы и характеристики.

Лексический анализатор построен с помощью инструмента flex [5]. Синтаксический анализатор построен с помощью инструмента GNU Bison [6]. В результате работы лексического и синтаксического анализатора на основе входного описания строится дерево абстрактного синтаксиса (AST). Анализатор ограничений принимает на вход AST, проверяет консистентность описаний (например, корректность описываемых в таблице характеристики или отсутствие атрибутов, которые противоречат друг другу) и строит внутреннее представление библиотеки. Внутреннее представление реализовано в виде наборов классов C++, а его построение и анализ на соответствие ограничениям реализовано в виде методов этих классов.

Трансляция RTL-моделей на языке Verilog

Трансляция Verilog-описаний (см. рис.1) рассматривается с двух сторон: трансляция в язык FIRRTL и во внутреннее представление САПР Utopia

EDA (Net). Net – это промежуточное представление уровня логических вентилях, поддерживающих иерархическое представление.

Транслятор реализован как программный компонент САПР Utopia EDA на языке C++. Основными компонентами транслятора являются:

- 1) транслятор Verilog-описаний в промежуточное представление;
- 2) транслятор в FIRRTL;
- 3) транслятор в Net.

Входом транслятора Verilog-описаний является файл описания функционального устройства на языке Verilog. При трансляции в Net компонент выступает в качестве фронтенда инструмента логического синтеза. При трансляции в FIRRTL результатом работы компонента является описание входного функционального устройства на языке FIRRTL.

FIRRTL предназначен преимущественно для моделей, подходящих для синтеза, и характеризуется более строгими правилами, чем Verilog, вследствие чего требуется проведение дополнительного анализа и преобразований для обеспечения корректной конвертации функционального устройства, описанного на Verilog, соответственно на язык FIRRTL. В связи с этим для разбора Verilog-описаний было решено использовать инструмент Yosys [9]. В рамках Yosys доступны необходимые инструменты и методы преобразований. Yosys транслирует Verilog-описание в собственное внутреннее представление — RTLIL. По своему устройству изначально оно близко к AST, но последовательное применение трансформаций позволяет снизить уровень абстракции модели до уровня логических вентилях, что значительно упрощает разработку транслятора в Net.

Транслятор в FIRRTL (FIRRTLPrinter) принимает на вход непосредственно RTLIL представление и в процессе работы получает описание исходной RTL-модели на языке FIRRTL. В основе процесса трансляции лежит рекурсивный обход RTLIL. Чтобы обеспечить корректность генерируемого описания, с помощью инструмента Yosys над промежуточным представлением совершается ряд трансформаций. В частности, основными являются преобразования в эквивалентное представление в виде графа, узлами которого являются допустимые для FIRRTL операции (add, sub, mux, pad, dshr, shr, dshl, shl и другие).

Транслятор в Net (NetBuilder) также принимает на вход RTLIL и также основан на его рекурсивном обходе. Основным отличием от транслятора в FIRRTL, помимо конструирования Net, является дополнительный набор трансформаций, которые совершаются над промежуточным представлением самим инструментом Yosys. Эти трансформации включают разбиение сущностей на отдельные биты, а также преобразование представления к эквивалентному виду без don't care битов.

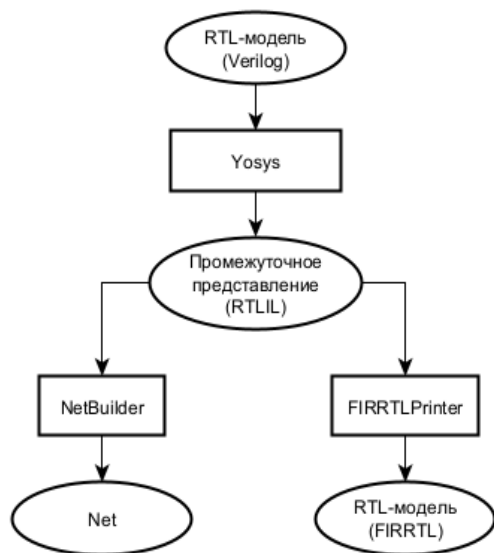


Рис.1. Реализованный процесс трансляции Verilog

Сравнение с существующими решениями

Ранее Synopsys предоставляла исходный код своей библиотеки для трансляции Liberty, но на данный момент компания закрыла к нему доступ и сохранилась только архивная версия [7], которая устарела и не соответствует текущей спецификации формата. Помимо этого, существуют открытые решения от сторонних разработчиков, например libertyParser [8], используемый на данный момент в Utopia EDA. Такие решения проводят только разбор синтаксиса, но не осуществляют анализ содержимого. К тому же возникает проблема интеграции: libertyParser написан на Python, в то время как разработка Utopia EDA ведется на C++. Разрабатываемое решение призвано решить эти проблемы.

Для синтаксического анализа Verilog-описаний существуют открытые инструменты (Slang [10], Surelog [11] и др.) с поддержкой стандартов Verilog (IEEE 1364) и SystemVerilog (IEEE 1800) [12]. Результатом трансляции, как правило, является AST. Однако, для построения корректного FIRRTL-описания недостаточно только AST – фактически, требуются элементы логического синтеза. С другой стороны, используемый инструмент Yosys способен транслировать Verilog описания функциональных устройств в FIRRTL, хотя его функциональность весьма ограничена: трансляция реализована только для комбинационных схем. Разрабатываемое решение позволяет осуществлять трансляцию более широкого подмножества Verilog.

Перспективы использования в процессе логического синтеза

Разработанные модули будут использоваться в Utopia EDA. Рассмотрим часть маршрута проектирования в САПР Utopia EDA совместно с разрабатываемыми подсистемами трансляции (см. рис.2).

Транслятор Verilog получает на вход файл с описанием RTL-модели на Verilog и осуществляет трансляцию данного описания во внутреннее пред-

ставление САПР (Net). Затем остальные компоненты САПР работают с полученным внутренним представлением. Внедрение инструмента позволит осуществить процесс логического синтеза с использованием RTL-моделей, написанных на Verilog, а также необходим для обеспечения совместимости с различными инструментами, такими как CIRCT. Кроме того, использование плагинов для Yosys позволит использовать такие HDL как SystemVerilog, например, с помощью плагина synlig [13], с относительно небольшими изменениями транслятора.

Транслятор Liberty принимает на вход файл с описанием технологической библиотеки в соответствующем формате. На выходе – структура данных, хранящая считанные сущности библиотеки и Net. Внедрение инструмента в САПР позволит осуществить процесс технологического отображения с использованием описаний технологических библиотек в формате Liberty.

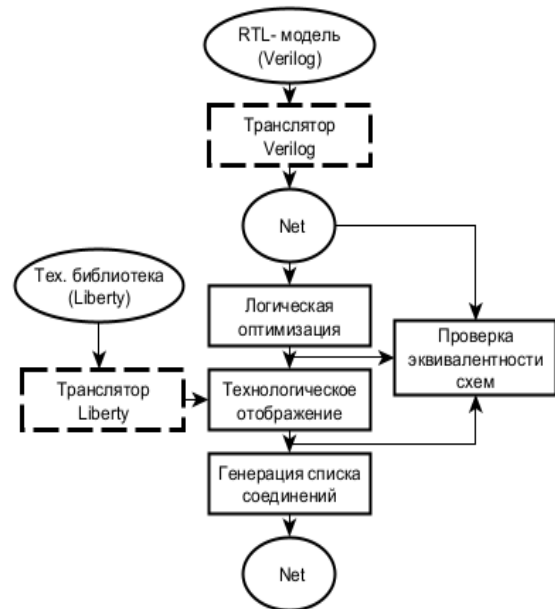


Рис.2. Схема части маршрута проектирования в САПР Utopia EDA (внедряемые модули отмечены пунктирной линией)

Заключение

Трансляция описаний RTL-моделей и моделей технологических библиотек является важной частью САПР. В рамках данной работы были реализованы прототипы компонентов трансляции: трансляторы Verilog и Liberty. Основными направлениями дальнейшей работы являются оптимизация, отладка, доработка разработанных прототипов и разработка интегрированного в Utopia EDA модуля для трансляции Liberty во внутреннее представление САПР (Net).

Список литературы

1. "The Institute of Electrical and Electronics Engineers IEEE Std 1364™-2005 (Revision of IEEE Std 1364-2001)", IEEE Standard for Verilog® Hardware Description Language, April 2006, ISBN 0-7381-4850-4.

2. Спецификация Liberty : [сайт]. – 2024. - URL: https://people.eecs.berkeley.edu/~alanmi/publications/other/liberty07_03.pdf (Дата обращения: 31.01.2024). – Текст : электронный.
3. Репозиторий CIRCT : [сайт]. – 2024. - URL: <https://github.com/circt> (Дата обращения: 31.10.2024). – Текст : электронный.
4. Репозиторий FIRRTL : [сайт]. – 2024. - URL: <https://github.com/chipsalliance/firrtl-spec/releases> (Дата обращения: 31.01.2024). – Текст : электронный.
5. Сайт flex : [сайт]. – 2024. - URL: <http://gnu.ist.utl.pt/software/flex/flex.html> (Дата обращения: 31.01.2024). – Текст : электронный.
6. Сайт GNU Bison : [сайт]. – 2024. - URL: <https://www.gnu.org/software/bison/> (Дата обращения: 31.01.2024). – Текст : электронный.
7. Неофициальный репозиторий парсера Liberty от Synopsys : [сайт]. – 2024. - URL: <https://github.com/eclufsc/libertyParser> (Дата обращения: 31.10.2024). – Текст : электронный.
8. Репозиторий Liberty Parser : [сайт]. – 2024. - URL: <https://codeberg.org/tok/liberty-parser> (Дата обращения: 31.10.2024). – Текст : электронный.
9. Репозиторий Yosys : [сайт]. – 2024. - URL: <https://github.com/YosysHQ/yosys> (Дата обращения: 31.10.2024). – Текст : электронный.
10. Репозиторий Slang : [сайт]. – 2024. - URL: <https://github.com/slang-project/SLang> (Дата обращения: 31.01.2024). – Текст : электронный.
11. Репозиторий Surelog : [сайт]. – 2024. - URL: <https://github.com/chipsalliance/Surelog> (Дата обращения: 31.01.2024). – Текст : электронный.
12. "The Institute of Electrical and Electronics Engineers IEEE Std 1800™-2009 (Revision of IEEE Std 1800-2009)", IEEE Standard for SystemVerilog® Unified Hardware Design, Specification, and Verification Language, December 2009, ISBN 978-0-7381-6130-3.
13. Репозиторий synlig : [сайт]. – 2024. - URL: <https://github.com/chipsalliance/synlig> (Дата обращения: 31.01.2024). – Текст : электронный.

СУШИЛЬНАЯ КАМЕРА, КАК СПОСОБ УМЕНЬШЕНИЯ БРАКА ПРИ ПЕЧАТИ НА 3D ПРИНТЕРЕ

*Цуркан А.А., Волошин А.Д.,
Клюев Н.С., Коробок М.А.*
*Национальный исследовательский университет
«Высшая школа экономики»
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе представлен анализ материалов, используемых для 3D печати методом послойного наложения (FDM) и ряд проблем, возникающих при ненадлежащем хранении филамента в условиях влажной среды. На основе данного анализа рассматривается спроектированное устройство для нивелирования данной проблемы.

Введение

Аддитивное производство, широко известное как 3D-печать – современная технология, характеризующаяся непрерывным развитием в рамках последних 10 лет, называемое одним из трендов Индустрии 4.0 [1]. Одним из распространенных методов аддитивного производства является технология FFF (Fused Filament Fabrication), основанная на последовательной укладке слоев с использованием экструдированного полимерного материала, называемого «нитью». Это позволяет быстро создавать объекты сложной геометрии [2], адаптируя их под конкретные применения множества отраслей [3], а современные исследования в области направлены на то, чтобы сделать 3D принтеры более надежными и удобными в повседневной эксплуатации [4].

Среднегодовой темп роста рынка аддитивных технологий превышает 25%, ожидается, что в течение следующих 5-10 лет его вклад в мировую экономику ежегодно будет составлять от \$200 до \$600 млрд [5]. Спрос в области аддитивных технологий, в свою очередь, влечет за собой требования к улучшению качества напечатанных моделей и минимизации брака.

Одной из причин возникновения отходов является неправильное хранение катушек с пластиком, обусловленное гигроскопичностью материала – способностью поглощать и удерживать влагу внутри. В связи с этим, для эффективного использования пластика в качестве печатного материала необходимо предварительно проводить его сушку [6]. Игнорирование данного процесса приводит к гидролизу, химическому разложению полимерных макромолекул под воздействием воды при высоких температурах. Во время термической обработки полимера вода испаряется и расширяется, формируя пузырьки в материале, что снижает как оптические характеристики изделия, так и межслойное сцепление, приводя к разделению слоев в процессе печати и к разрушению конструкции изделия во время эксплуатации [7]. Таким образом, целью данной работы стала разработка сушильной камеры для филамента.

При обзоре аналогов было выделено 3 основные группы сушильных камер: с сушкой с помощью ламп накаливания (дешевый ценовой сегмент), сушка с помощью РТС нагревателя (средний ценовой сегмент) и сушка с помощью вакуума (производственные сушильные камеры). Данные устройства объединяет отсутствие заготовленных профилей для сушки филаментов разного типа и вместимость, которая не превосходит 3ех катушек пластика.

Оценка гигроскопичности филамента

Для проведения эксперимента были выбраны образцы PLA и ABS, как самые часто используемые полимеры в процессе 3D печати. Перед началом печати образцы филамента хранились в сухих условиях, с контролируемой влажностью, в течение 48 часов для стабилизации исходного состояния, после чего были помещены в среду с влажностью 100% на периоды в 30 и 60 минут для имитации различных условий хранения и процента впитанной влаги. После воздействия влажной среды образцы использова-

лись для печати стандартизированных тестовых моделей (рис.1) на 3D-принтере, работающем по технологии FFF. Параметры печати (температура экструдера, скорость печати, толщина слоя) были стандартизированы для всех тестов.

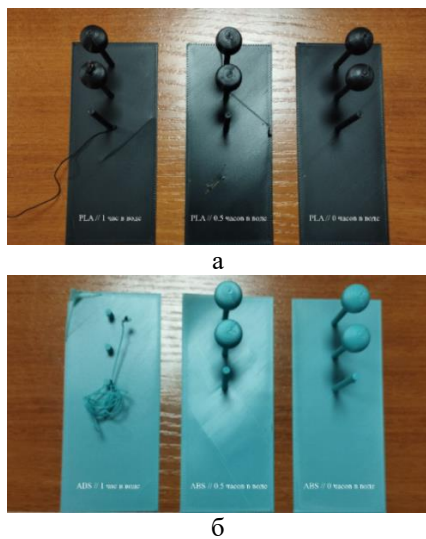


Рис.1. Сравнение качества напечатанных изделий из филамента с разными условиями хранения, PLA (а) и ABS (б)

Качество печатных изделий оценивалось визуально, с использованием методов поверхностной микроскопии для идентификации дефектов, связанных с влажностью. Из сравнения тестовых моделей очевидно влияние влажности пластика на качество печати. Гигроскопичность ABS больше гигроскопичности PLA, поэтому дефекты печати на тестовых моделях из ABS проявились значительно сильнее. Стоит отметить, что при печати проявились дефекты разных категорий. Например, при печати ABS пластиком различной влажности, можно выявить недостаточную адгезию между слоями модели, в то время как на дефекты на PLA тестах принимали вид «спагетти» эффекта и неровностей напечатанных слоев.

Проводя дополнительный анализ температурных режимов и параметров сушки филаментов ряда производителей, была сформирована таблица средних значений (таблица 1), по которой можно удостовериться, что не имеет смысла в превышении конечным устройством стабильной температуры сушки в +110°C. Данный вывод позволяет задать стабильный и оптимальный уровень мощности и потребления энергии вне зависимости от типа пластика.

Таблица 1. Сравнение температурных режимов сушки

| Материал | Температура, °C | Время обработки, ч | Гигроскопичность, % |
|-----------|-----------------|--------------------|---------------------|
| PLA | + 53 | >4 | 2 – 50 |
| PVA | + 45 | >4 | 5.6 |
| Neylon | + 103 | >4 | 3.09 |
| Eternal | + 60 | >4 | 2.3 |
| Easy Flex | + 50 | >4 | 2.1 |

| | | | |
|------|------|----|---------|
| ABS | + 80 | >2 | 0.2-0.3 |
| SBS | + 65 | >4 | 0.06 |
| PETG | + 70 | >5 | 0 |
| HIPS | + 55 | >6 | 0 |

Схемотехническое проектирование

Для управления процессом сушки пластика была разработана печатная плата на основе микроконтроллера AVR ATmega. Операционное программное обеспечение для управления элементами сушилки, включая нагревательные элементы и сбор данных с термодатчиков, максимально занимает 13Кб памяти, что подтвердило целесообразность выбора данного микроконтроллера.

Электропитание контроллера обеспечивается от источника 5 В с использованием фильтрующих конденсаторов для стабилизации напряжения. Частота работы микроконтроллера задается внешним кварцевым резонатором 12 МГц.

Точность АЦП достигается за счет подачи на пин AREF напряжения 5 В с дополнительным фильтрующим конденсатором, шаг аналого-цифрового преобразователя при этом составляет 0.0049 В. Программирование микроконтроллера производится через интерфейс AVR ICSP с использованием стандартного программатора.

Электропитание платы осуществляется от сети 220 В переменного тока. Для преобразования переменного тока в постоянный напряжения 12 В на плате установлен AC-DC преобразователь питания HS12P36SR. На плате также реализованы DC-DC преобразователи L78L05ABUTR и XC6206P332MR для понижения напряжения до 5 В и 3.3 В соответственно, необходимых для питания микроконтроллера, термисторов и ряда других компонентов.

Управление питанием вентиляторов осуществляется через оптопару TLP627 с транзисторным выходом, при подаче питания через клеммник ХН-2pin.

В свою очередь, регулировка мощности нагревательных элементов достигается путем симисторного диммирования, для чего применяются оптопары МОС-3020 и РС814Х1СS29F.

Пользовательский интерфейс реализован с помощью дисплея Nexion ENHANCED NX4024K032, подключаемого через UART, что облегчает взаимодействие с устройством.

Полная принципиальная схема изделия представлена на рис.2.

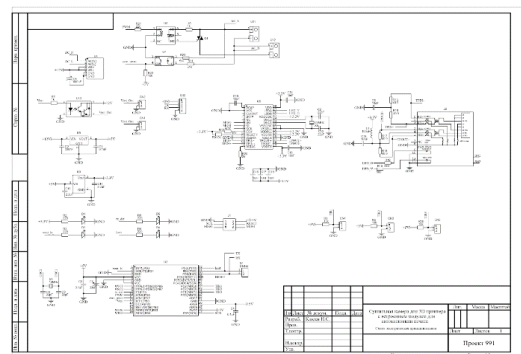


Рис.2. Принципиальная схема сушильной камеры

На рис.3 изображена разведенная в EasyEDA печатная плата с предусмотренными монтажными отверстиями диаметром 3 мм и изоляцией зоны высокого напряжения. Габариты платы 90x90 мм.

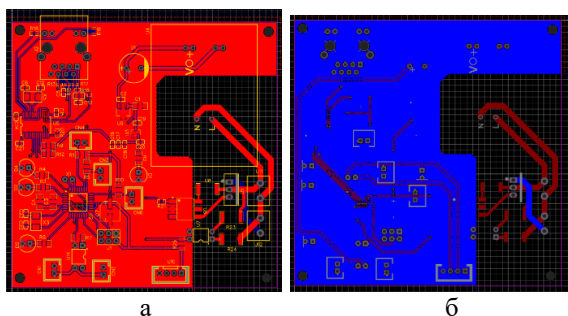


Рис.3. Передняя(а) и оборотная (б) сторона печатной платы

Проектирование корпуса устройства

Проектирование устройства выполнялось в среде SolidWorks, где прототип устройства выполнен в виде одного сборочного объекта (рис.4), совмещающего сушильную камеру и отсек электроники. Подобный форм-фактор упрощает производство деталей и увеличивает конструкционную прочность всей сборки. Материалом изготовления корпуса была выбрана листовая сталь из-за прочностных характеристик, дешевизны и простоты обработки.

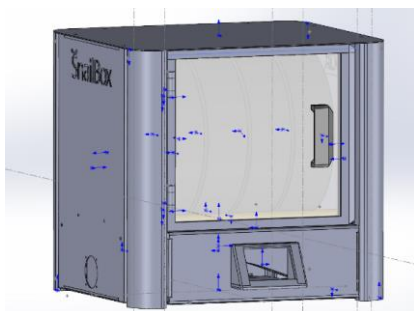


Рис.4. 3D модель сушильной камеры

Сушильная камера изготовлена в виде параллелепипеда с закругленными ребрами. Четыре угловых ребра являются как каркасом конструкции, к которому крепятся листовые детали, так и выполняют функционал скрытия проводки датчиков, необходимых для размещения внутри камеры. Такая конструкция является легкой в сборке и прочной. Оргстекло, из которого выполнена передняя грань позволяет пользователю контролировать процесс сушки с возможностью аварийного отключения из сети при возникновении неполадок.

Все детали скрепляются между собой винтами, что делает возможным при необходимости полностью разобрать камеру. Катушки пластика внутри располагаются вертикально, что позволяет сушить до 5 катушек филамента параллельно и обеспечить одновременный доступ к любой из них. Сами «намотки» располагаются на гладких цилиндрических ползьях, что делает возможным подачу пластика без

использования дополнительных двигателей, обеспечивающих размотку катушки, а также предотвращает крен катушек во время подачи полимера.

Трубка забора и создания воздушного потока предусматривает установку двух вентиляторов, для увеличения воздушного потока, проходящего сквозь нагреватель. Вся система подачи воздуха полностью изолирована от камеры электроники, для того чтобы не допустить её нагрева в процессе эксплуатации.

В ходе теплового моделирования (рис.5) было выявлено, что дополнительные стенки, выполняющие роль термокамеры, улучшают энергоэффективность устройства, уменьшают нагрев внешних стенок камеры, что уменьшает травмоопасность, а также защищают опоры катушек от провисания. Реализация такой модификации является перспективным направлением работы.

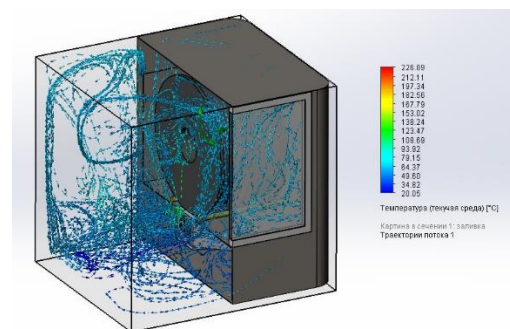


Рис.5. Температурное моделирование для корпуса сушильной камеры

Таким образом было спроектировано и собран полнофункциональный образец устройства (рис.6).

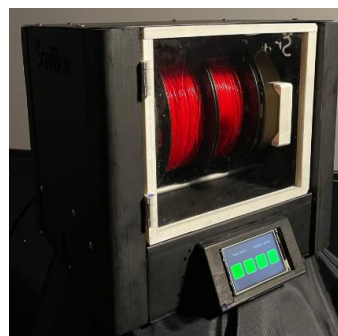


Рис.6. Сборка сушильной камеры

Заключение

Результатом разработки является исследование свойств различных типов филамента, а также устройство для их сушки перед печатью на 3D-принтере.

Все управляющие компоненты смонтированы на печатной плате и изолированы в корпусе под основным отсеком для печати, камера нагревания закрыта защитным корпусом. При работе с устройством пользователи самостоятельно могут регулировать температуру сушки, а также следить за влажностью и температурой с помощью показателей на экране.

Список литературы

1. Prashar G., Vasudev H., Bhuddhi D. Additive manufacturing: Expanding 3D printing horizon in industry 4.0 // Int. J. Interact. Des. Manuf. (IJIDeM). – 2022. – V.17. – P.1-15.
2. Singh S., Singh G., Prakash C., Ramakrishna S. Current status and future directions of fused filament fabrication // J. Manuf. Process. – 2020. – V.55. – P.288-306.
3. Dey A., Roan Eagle I.N., Yodo N. A Review on Filament Materials for Fused Filament Fabrication // J. Manuf. Mater. Process. – 2021. – V.5. – P.69.
4. Бородин Ю. Н., Сбоева И. А. Техпроцесс 3Д-печати // Научно-образовательный журнал для студентов и преподавателей «StudNet». – 2021. – № 1. – С.2-3.
5. Голубничая Я.Р., Проскуряков Н.Е. Перспективные технологии 3Д-печати // Известия ТулГУ. Технические науки – 2017. – №.9. – С.5-6.
6. Chabaud G., Castro, M., Denoual, C., Le Duigou A., Hygromechanical properties of 3D printed continuous carbon and glass fibre reinforced polyamide composite for outdoor structural applications // Addit. Manuf. – 2019. – V.26. – P.94-105.
7. Zaldivar R.J., McLouth T.D., Ferrelli G.L., Patel D.N., Hopkins A.R., Witkin D., Effect of initial filament moisture content on the microstructure and mechanical performance of ULTEM® 9085 3D printed parts. // Addit. Manuf. – 2018. – V.24. – P.457-466.

МЕТОДИКА ОПРЕДЕЛЕНИЯ БАЗОВЫХ ПОКАЗАТЕЛЕЙ НАДЕЖНОСТИ ЭКБ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Костиков Д.В., Кацнельсон А.И.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Научно-исследовательская работа посвящена новой методике оценивания надежности электронных компонентов, в особенности полупроводниковых приборов, с использованием методов машинного обучения для прогнозирования показателей безотказности. Это имеет практическое значение для производителей электронных средств, помогая им улучшить качество прогнозирования надежности своих изделий и повысить уровень технического развития и конкурентоспособности на рынке.

Введение

Оценка надежности деталей и устройств - ключевой аспект в инженерной практике. Анализ надежности позволяет выявить причины отказов и увеличить срок службы электронных компонентов [1]. Для этого проводится классификация детали по вероятности отказа, однако доступные справочники по надежности устарели и неточны. Решение - математическая

модель, предсказывающая надежность компонента. Цель проекта - создание автоматизированной системы для определения отказоустойчивости на основе характеристик электронных компонентов, включая данные из справочника по надежности [2]. В работе рассматриваются диоды.

Общая теория надежности

Сейчас надежность электронной компонентной базы (ЭКБ) находится на очень высоком уровне и основной метрикой для её оценки является интенсивность отказов, или же Failure Rate. С помощью формулы это можно выразить так:

$$\lambda_i^* = \frac{n(\Delta t_i)}{N_{cpi} \Delta t_i}$$

где $n(\Delta t_i)$ — это количество элементов, которые получили отказ на некоем временном интервале. N_{cpi} принимается за среднее количество элементов, работающих без отказов на некоем временном интервале. И последнее, Δt_i , является шириной некоего временного интервала.

Основные три фактора, используемые в математических моделях для расчета интенсивности отказов, это факторы, влияющие на отказы кристалла, факторы влияющие на отказы корпуса и устойчивость к электростатическому заряду.

Для расчета используется значение эксплуатационной интенсивности отказа для групп, считается это по данной формуле:

$$\lambda_{\Sigma} = \lambda_{\text{Б}} \prod_{i=1}^m K_i$$

где $\lambda_{\text{Б}}$ является показателем базовой интенсивности отказов для изделий данной группы;

K_i — это коэффициенты, которые непосредственно влияют на итоговую эксплуатационную интенсивность отказов, отражающие разного рода факторы; m — это число учитываемых факторов.

Индивидуальные или же базовые интенсивности отказов $\lambda_{\text{Б}}$ обычно берутся из справочников, либо же подсчитываются исходя из сложных моделей.

Предлагаемое решение

Работа предлагает новый подход, который автоматизирует процесс формирования справочников надежности электронных компонентов и сокращает временные, материальные и финансовые затраты [3]. Он включает в себя использование методов машинного обучения на этапе испытаний новых и перепроверки серийных полупроводниковых изделий или других типов электрорадиоизделий в организациях разработчиках и производителях ЭКБ, а также использование данных непосредственно разработчиками электронных средств [4]. Оценка проводится для вновь разрабатываемых компонентов или по ускоренному формату, или по предиктивному анализу с использованием методов машинного обучения, или в смешанном формате, совмещая методы машинного обучения и испытания ЭКБ.

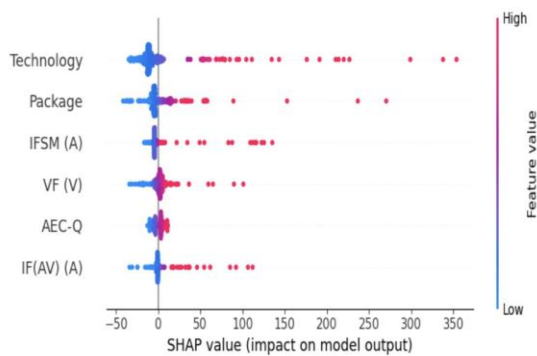


Рис. 4. SHAP-значение при прогнозировании показателей надежности диодов

Градиентный бустинг — метод машинного обучения для построения ансамблей моделей, объединяющих слабые модели, чаще всего линейные или логистические регрессии [8]. Он базируется на пошаговом улучшении модели, начиная с базовой, и на каждом шаге строится новая модель для корректировки ошибок предыдущей.

Функция потерь определяет ошибку модели в прогнозировании целевой переменной. Градиентный бустинг минимизирует эту функцию, вычисляя градиент функции потерь относительно предсказаний модели. Новая модель строится так, чтобы предсказывать целевую переменную, улучшая предсказания ансамбля. Процесс повторяется до сходимости или достижения порога.

Градиентный бустинг имеет различные вариации, такие как градиентный бустинг с использованием деревьев решений, XGBoost, LightGBM и CatBoost. Каждая из них вносит свои оптимизации и улучшения.

Результаты предсказаний с использованием CatBoost показали MAPE = 24%, что значительно уступает случайному лесу.

Результаты

Лучшие результаты были достигнуты с использованием алгоритма случайного леса, который дал MAPE (среднюю абсолютную процентную ошибку) на уровне 2%. Это превосходит результаты градиентного бустинга и линейной регрессии. Хотя обычно модель градиентного бустинга показывает более высокое качество, в данной задаче прогнозирования надежности базовых показателей ситуация обратная. Основными параметрами для прогнозирования надежности диодов технологической группы оказались технология изготовления, корпус, максимальный импульсный ток диода (IFSM), падение напряжения на переходе (VF), средний ток диода (IF), и сертификация (AEC-Q). Учет только этих параметров снизил точность прогноза надежности всего на 0.4%, как видно из таблицы 1.

Таблица 1. Сравнение моделей

| Линейная регрессия | Случайный лес | Градиентный бустинг |
|--------------------|---------------|---------------------|
| 96608893193% | 24% | 2% |

Заключение

Исходя из наших исследований, можно сделать следующие выводы:

1. Надежность электронных компонентов играет важную роль в обеспечении долгосрочной работы электронных устройств.

2. Существующие справочники по надежности устарели и не подходят для современных задач, что может привести к большим ошибкам в прогнозировании.

3. Создание новых справочников требует больших затрат и все равно может не обеспечить точность в прогнозах надежности.

4. Исследование различных методов машинного обучения показало, что случайный лес с 2% средней абсолютной процентной ошибкой является наилучшим вариантом.

5. Градиентный бустинг, несмотря на более высокий уровень ошибки в 24%, также пригоден для более точных прогнозов, хотя и уступает алгоритму случайный лес в показателе точности.

6. Дальнейшие исследования могут сосредоточиться на улучшении предсказаний надежности с использованием более сложных методов машинного обучения и расширении базы данных для различных классов компонентов.

В итоге, разработка математических моделей с применением методов машинного обучения, таких как случайный лес и градиентный бустинг, представляется перспективной для оценки электронных компонентов, что способствует качественной и долгосрочной эксплуатации технических систем.

Список литературы

1. Расчет надежности радиоэлектронной аппаратуры // ncfu.ru URL: https://www.ncfu.ru/export/uploads/imported-from-dle/op/doclinks2015/Metod_Diagn.-i-nad.-el.teh.-ust.-PZ_13.03.02_12.05.2015.pdf (дата обращения: 12.10.2023).
2. Надёжность технических систем: справочник / Ю. К. Беляев [и др.] ; под ред. И. А. Ушакова. – М.: Радио и связь, 1985. – 608 с.
3. Черкесов Г.Н. Оценка надежности систем с учетом ЗИП. - СПб.: БХВ-Петербург, 2012. - 478 с.
4. Оценка надежности городской водопроводной сети при дефиците информации методом искусственных нейронных сетей // urfu.ru URL: https://elar.urfu.ru/bitstream/10995/121644/1/978-5-7996-3616-6_2023.pdf (дата обращения: 12.10.2023).
5. Taiwan Semiconductor // taiwansemi.com URL: <https://www.taiwansemi.com/en/> (дата обращения: 12.10.2023).
6. Texas Instruments // ti.com // URL: <https://www.ti.com/> (дата обращения: 12.10.2023).
7. Решающие деревья // Учебник по машинному обучению URL: <https://academy.yandex.ru/handbook/ml/article/reshayus-hchiye-derevya> (дата обращения: 12.10.2023).
8. Gradient boosting: Distance to target // explained.ai URL: <https://explained.ai/gradient-boosting/L2-loss.html> (дата обращения: 12.10.2023).

ОБЗОР ТЕХНОЛОГИЙ EDGE AI, ПРИМЕНЯЕМЫХ В АВТОИНДУСТРИИ

Щукин А.А.¹, Рожин А.К.²

Национальный исследовательский университет
«Высшая школа экономики»,
¹департамент программной инженерии,
²департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Рассмотрены различные типы нейронных сетей в задачах классификации, сегментации и обнаружения, а также проведено их сравнение в производительности и наборе операций для последующей реализации этих сетей на нейроускорителе.

Введение

Нейронные сети с большой точностью зачастую имеют большое количество параметров, что влечет за собой вычислительную сложность, как для обучения, так и для запуска нейронной сети, однако иметь высокопроизводительную систему на автотранспорте – экономически неэффективно. Поэтому, встает задача выбора вычислительно-эффективной модели классификации или сегментации, но, наряду с этим, модель должна обладать достаточной точностью для использования в надежных системах ADAS [1].

В данной работе проведен сравнительный анализ некоторых open-source моделей в области edge AI.

В ходе анализа рассматривались следующие метрики нейронных сетей: количество MAC операций, количество параметров, вес модели в байтах, наличие Pooling операций, а также наличие операций, не поддерживаемых систолическим массивом. В каждом разделе представлено лишь 3 модели, не включены модели, имеющие большой размер, маленькую точность и/или большое количество MAC операций.

Подсчеты производились с помощью утилиты onnx_tool для языка Python3.

Модели в области 2D входных данных

В таблице 1 представлено сравнение моделей для задачи классификации. Из особенностей можно выделить несколько:

1. Всем моделям присуще использование пулинга.
2. Использование линейных и квадратичных функций активации в модели MobileNetV3_large вместо нелинейной сигмоиды позволило обойтись меньшими вычислительными затратами, по сравнению с остальными моделями, при этом сохраняя сопоставимую точность.

Таблица 1. Модели задачи классификации в 2D

| Название модели | SqueezeNet [2] | MobileNetV3 large [3] | EfficientNetB0 [4] |
|-----------------|----------------|-----------------------|--------------------|
| Mac-операции | 8,3E8 | 2,32E8 | 6,01E8 |

| | | | |
|----------------------------|-------------------------|---------------------------------|----------------|
| Количество параметров, шт. | 1,25E6 | 5,45E6 | 5,2E6 |
| Память, Б | 5,28E7 | 5,89E | 1,18E7 |
| Top-1 ImageNet Accuracy | 57,5% | 75,2% | 77,5% |
| Pooling? | GlobalAvg Pool, MaxPool | GlobalAvg Pool | GlobalAvg Pool |
| Сложные операции | Flatten | Hardswish, HardSigmoid, Flatten | - |

Далее был проведен сравнительный анализ моделей задачи сегментации (Таблицы 2 и 3). Данная задача является более вычислительно сложной и модели обладают большим количеством MAC-операций. Сравнительная таблица представлена ниже.

Таблица 2. Модели класса сегментации

| Название модели | YOLOv5 n-seg | Segformer (MiT-b0) | VAN-tiny |
|----------------------------|---------------|--------------------|-----------------|
| Mac-операции | 4,1E9 | 5,41E8 | 9,62E9 |
| Количество параметров, шт. | 2E6 | 3,32E6 | 3,85E6 |
| Память(Б) | 3,08E8 | 1,5E8 | 1,92E8 |
| z-index | -0,417 (COCO) | -0,332 (ADE20k) | -0,304 (ADE20k) |

Таблица 3. Модели класса сегментации (продолжение)

| Название модели | YOLOv5s -seg | VAN-small |
|----------------------------|---------------|-----------------|
| Mac-операции | 1,42E10 | 2,6E9 |
| Количество параметров, шт. | 7,61E6 | 1,34E7 |
| Память, Б | 5,29E8 | 3,06E8 |
| z-index | -0,242 (COCO) | -0,189 (ADE20k) |

Ввиду отсутствия вычислительных мощностей для проверки моделей на едином датасете, вместо их реального значения используется z-index по датасету, на котором у них известна эффективность/точность.

В данной таблице не указаны сложные операции и наличие пулингов, но:

1. Семейство VAN [5] не имеет пулингов и имеет операции ReduceMean и Erf.

2. Семейство YOLOv5 [6] имеет только MaxPool, а среди операций присутствуют Where и Expand.

3. Семейство SegFormer [7] не обладает пулингами и имеет такие же операции, как и VAN модели.

Заметим, что добавление операций Where и Expand сильно ухудшило показатели производительности моделей семейства YOLO, при этом не привнося сильного улучшения производительности и размера модели. Ухудшение вызвано невозможностью параллелизации операции Where внутри одного слоя нейросети.

В таблице ниже представлены результаты сравнительного анализа моделей класса обнаружения. В данном классе еще меньше моделей ввиду еще большей вычислительной сложности данной задачи.

В таблице не указаны столбцы операций и пулингов, так как пулинги совпадают (MaxPool), операции тоже не отличаются (сложной является только Expand).

Таблица 4. Модели Класса обнаружения

| Название модели | Yolov5-s | Yolox-tiny | Yolov5-n | Yolox-nano |
|----------------------------|----------|------------|----------|------------|
| Мас-операции | 9,1E9 | 3,5E9 | 2,7E9 | 1,7E9 |
| Количество параметров, шт. | 7,2E6 | 5E6 | 1,9E6 | 2,24E6 |
| Память, Б | 4,8E8 | 1,58E8 | 2,5E8 | 1,03E8 |
| mAP score | 37,4 | 32,8 | 28 | 25,8 |

В результате сравнительного анализа данных моделей выявлено следующее:

1. При затратах на память вдвое или больше раз у моделей семейства YOLOv5, чем у семейства YOLOX [8], выходная производительность все равно на сопоставимом уровне.

2. Полученное количество очков на датасете пропорционально логарифму количества MAC-операций модели (с коэффициентом $\sim 7,05$)

Список литературы

1. Jiménez, F. Advanced Driver Assistance System for Road Environments to Improve Safety and Efficiency / F. Jiménez, J.E. Naranjo, J.J. Anaya, F. García, A. Ponz, J.M. Armingol // Transportation Research Procedia. — 2016. — Vol. 14. — P. 2245-2254.

2. Iandola, F.N. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size / F.N. Iandola, M.W. Moskewicz, K. Ashraf, S. Han, W.J. Dally, K. Keutzer // arXiv preprint arXiv:1602.07360. — 2016.

3. Howard, A.G. Searching for MobileNetV3 / A.G. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q.V. Le, H. Adam // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). — 2019. — P. 1314-1324.

4. Tan, M. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks / M. Tan, Q.V. Le // arXiv preprint arXiv:1905.11946. — 2019.

5. Guo, M.-H. Visual attention network / M.-H. Guo, C. Lu, Z.-N. Liu, M.-M. Cheng, S. Hu // Computational Visual Media. — 2022. — Vol. 9. — P. 733-752.

6. Ultralytics. YOLOv5: A state-of-the-art real-time object detection system // Ultralytics Docs. — 2021.

7. Xie, E. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers / E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo // Neural Information Processing Systems. — 2021.

8. Ge, Z. YOLOX: Exceeding YOLO Series in 2021 / Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun // arXiv preprint arXiv:2107.08430. — 2021.

КОММУНИКАЦИЯ ПРОЦЕССОРОВ В ГЕТЕРОГЕННОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЕ

Губазов Д.Д., Кателкин Н.И.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В докладе приведено описание ряда программно-аппаратных решений, обеспечивающих коммуникацию двух процессоров в вычислительной системе. Отражены наиболее значимые этапы реализации обмена данными в вычислительной системе.

Введение

Любая вычислительная система нуждается в программно-аппаратном комплексе средств, реализующем обмен данными между элементами системы, в частности между процессорами.

Техническая задача: обеспечить обмен данными между отладочной платой ПЛИС Kintex-7 и персональным компьютером любого вида посредством внешних интерфейсов и компьютерных шин.

Цель: объединение персонального компьютера и отладочной платы в единую вычислительную систему для тестирования ПО отладочной платы и совместной обработки данных в рамках проекта по созданию служебного устройства, предназначенного для ускоренного исполнения моделей машинного обучения (далее - "нейроускоритель").

Комплекс программно-аппаратных средств создан с использованием сред разработки аппаратного обеспечения (Vivado), инструментов разработки ПО (Vitis SDK, GCC, Zephyr OS) и языков программирования (C, C++).

Результатом работы считается успешная реализация обмена данными между платой ПЛИС и ПК по выбранному в ходе работы интерфейсу.

Решение

Конфигурация ПЛИС, включающая аппаратный блок интерфейса, и программное обеспечение, реализующее обмен данными. Конфигурация разработана в Vivado с использованием как готовых, так и пользовательских аппаратных блоков.

Конфигурация ПЛИС и отладочная плата составляют *систему на базе процессора RISC-V*.

Программное обеспечение написано на языке программирования C/C++ и использует открытую операционную систему Zephyr OS для запуска на отладочной плате.

Система на базе RISC-V с реализованной аппаратной поддержкой выбранного в ходе работы интерфейса является *основой нейроускорителя*.

Обмен данными между ПК и нейроускорителем заключается в передаче по внешнему интерфейсу моделей машинного обучения для последующего инференса.

ПК и нейроускоритель образуют единую вычислительную систему с ускоренным исполнением моделей машинного обучения.

Описание вычислительной системы. Выбор интерфейса

Вычислительная система включает в себя персональный компьютер и отладочную плату ПЛИС Kintex-7.

Персональный компьютер может иметь любой вид (настольный, ноутбук, неттоп и др.).

Отладочная плата Kintex-7 поддерживает на физическом уровне наиболее популярные интерфейсы сопряжения с компьютером (USB, PCI Express, Ethernet и т.д.).

Для решения поставленной задачи необходимо использовать интерфейс, отвечающий следующим требованиям:

1. Поддерживается на физическом уровне отладочной платой Kintex-7 и персональным компьютером любого вида.

2. Использует наиболее простой механизм сопряжения устройств.

USB версии 2.0 [1] – единственный интерфейс, отвечающий всем указанным требованиям.

Элементы вычислительной системы будут соединены посредством USB.

Особенности аппаратной реализации USB

Аппаратный блок USB настраивается и включается в конфигурацию ПЛИС. После генерации обновленной конфигурации следует тестирование блока - завершающий этап аппаратной реализации USB.

Для тестирования блока используется следующая конфигурация ПЛИС:

1. Soft-процессор MicroBlaze с поддержкой AXI.
2. Локальная BRAM память.
3. Коммутатор AXI-устройств (AXI Interconnect).
4. Контроллер прерываний.
5. Контроллер внешней DDR3 оперативной памяти.
6. Стандартный таймер и блок сброса системы.

Все указанные выше периферийные устройства поддерживают AXI и используют стандартный механизм подключения к шине.

Данная конфигурация используется в связи с уникальными преимуществами soft-процессора MicroBlaze, значительно упрощающими тестирование аппаратных блоков:

1. Программная библиотека для взаимодействия с IP-блоками Vivado.

2. Интегрированная с Vivado среда разработки ПО Vitis IDE.

После успешного прохождения тестирования происходит интеграция аппаратного блока USB в систему на базе процессора RISC-V с последующей разработкой ПО для обмена данными с ПК.

Изучение особенностей отладочной платы ПЛИС Kintex-7

Плата использует USB-конвертер CP2102, непосредственно соединенный с ПЛИС. Данные, передаваемые по USB отладочной плате, конвертер представляет в виде потока байт UART [2]. На выходных контактах конвертера с помощью подтягивающих резисторов установлено напряжение +3.3В.

Реализация аппаратной поддержки USB на ПЛИС

Реализация поддержки USB заключается в создании или использовании готового аппаратного блока, обеспечивающего системе на ПЛИС возможность обмена данными с внешними USB устройствами, и получения обновленной конфигурации ПЛИС, поддерживающей внешний USB-конвертер CP2102 отладочной платы.

Аппаратный блок должен отвечать следующий требованиям:

1. Поддержка AXI для сопряжения с процессором.

2. Поддержка UART для сопряжения с конвертером CP2102.

3. Поддержка прерываний и совместимость с контроллером прерываний системы на ПЛИС.

В результате анализа готовых решений найдены аппаратные блоки, полностью удовлетворяющие указанным выше требованиям [3].

Интеграция аппаратного блока AXI UART Lite в конфигурацию ПЛИС состоит из нескольких этапов:

1. Подключение к интерфейсу AXI с помощью коммутатора.

2. Подключение к контроллеру прерываний.

3. Создание constraints-файла [4] (сопряжения с выходными контактами CP2102).

4. Получение диапазона адресов в адресном пространстве системы.

5. Синтез, имплементация.

6. Запись полученной конфигурации в ПЛИС.

Сопряжение с процессором на ПЛИС

Блок не может обращаться к другим периферийным устройствам, следовательно, подключается к коммутатору AXI как slave-устройство. Выходной контакт генератора аппаратных прерываний соединяется с контроллером прерываний системы.

Соединение блока с конвертером

USB-конвертер является внешним по отношению к ПЛИС устройством. Для сопоставления выходных контактов аппаратного блока и конвертера используются constraint-файлы. Файл состоит из двух основных групп команд:

- Первая группа команд устанавливает соответствие между логическими сигналами на схеме с физическими контактами ПЛИС. Выходы аппаратного блока сопоставляются с контактами ПЛИС, к которым подключен конвертер.

- Вторая группа команд устанавливает напряжение, соответствующее логической единице. Логической единице конвертера соответствует напряжение +3.3В, следовательно, для каждого контакта ПЛИС логическая единица также соответствует напряжению +3.3В.

Получение диапазона адресов в адресном пространстве системы

Аппаратный блок не требует большого диапазона адресов, следовательно, может получить свободный диапазон адресов любого разрешаемого средой разработки размера.

Синтез, имплементация, генерация обновленной конфигурации выполнены без ошибок.

Следующий этап — программное тестирование блока.

Программное тестирование аппаратного блока USB

Для создания тестирующей программы будем использовать библиотеки для работы с прерываниями и USB UART блоком. Программа написана на языке C и содержит функции отправки, приема произвольного числа байт по UART, и их обработки.

Конфигурация ПЛИС считается подходящей для решения поставленной задачи, если в результате работы программы выполнены следующие требования:

1. Система успешно отправляет произвольный объем данных на ПК.
2. Система корректно принимает данные с ПК.
3. Прием данных происходит в момент их поступления независимо от загруженности процессора.

Требование (3) необходимо для определения правильного взаимодействия блока с контроллером прерываний.

В результате работы программы требования выполнены. Аппаратный блок успешно добавлен в систему.

Интеграция в систему на базе RISC-V

Последний этап разработки аппаратного обеспечения - интеграция в исходную систему на базе процессора RISC-V.

RISC-V использует встроенный контроллер прерываний IPIC.

Единственная вследствие сходства тестирующей и исходной систем задача данного этапа - подключение аппаратного блока USB к контроллеру прерываний IPIC процессора RISC-V и получение итоговой конфигурации ПЛИС.

Конфигурация обновленной RISC-V системы получена успешно.

Дальнейшие шаги по достижению поставленной цели заключаются в разработке программного обеспечения. Разработка аппаратного обеспечения завершена.

Протокол взаимодействия устройств поверх физического уровня

Протокол взаимодействия нацелен на динамическую загрузку/подгрузку моделей машинного обучения в память фирменного нейроускорителя, ускоряющего инференс моделей, и последующее произвольное исполнение этих моделей. Модель передачи данных можно разделить на 3 уровня согласно таблице 1.

На прикладном уровне для передачи структурированных сообщений использовалась библиотека FlatBuffers – портативная, минималистичная и удобная для наших целей. В основе FlatBuffers лежат *схемы* - файлы, которые напоминают программный код, но не содержат исполняемого кода как такового. В данном случае они описывают структуру, согласно которой наши сообщения сериализуются в массивы байтов и десериализуются из них, т.е. какие поля содержатся в сообщении, какое предназначение и какой тип данных они имеют и в каком порядке они идут.

Таблица 1. Уровни взаимодействия между двумя устройствами

| | | | |
|------------------------------|---|---------------------------------------|-----------------------------|
| Прикладной уровень | Структурированные сообщения в FlatBuffers-формате (одного из 6 типов) | | |
| Транспортный уровень | Сигнатура сообщения (32-bit) | Длина сообщения в байтах (32-bit, LE) | Полезная нагрузка сообщения |
| Физический/канальный уровень | USB, UART, PCI Express | | |

Механизм загрузки моделей на плату

Перед тем, как начать передачу данных модели, устройство-хост (персональный компьютер) посылает на устройство-ускоритель (отладочную плату) сообщение типа ListInstalledOpsRequest. ПО, установленное на плате, обязано ответить на данное сообщение другим сообщением типа ListInstalledOpsResponse, содержащим список 128-битных идентификаторов (GUID) моделей, на данный момент загруженных в память платы.

ПО устройства-хоста сверяет список этот идентификаторов со списком идентификаторов моделей, которые требуется загрузить на плату, и в итоге загружает только те модели, которые ещё не загружены. Отправляется сообщение типа InstallOpRequest, содержащего в себе идентификатор модели, пара-

метры модели, её ядро (скомпилированный программный код модели; нами использовался формат исполняемых файлов ELF), размерности и количество тензоров, принимаемых на вход и подаваемых на выход. Устройство-ускоритель отвечает на это сообщением `InstallOpResponse`, либо подтверждающим успешность передачи модели, либо указывающим на ошибку при передаче модели.

Механизм запуска моделей на плате

Чтобы начать исполнение модели, хост отправляет сообщение типа `InvokeOpRequest`, содержащее список тензоров, над которыми требуется провести вычисления.

Устройство-ускоритель проводит вычисления и отправляет сообщение `InvokeOpResponse`, при ошибке содержащее пустой список тензоров, а при успешном завершении вычислений – список выходных тензоров.

Описание ПО для устройства-хоста

Программное обеспечение для устройства-хоста оформлено в виде C++-библиотеки, выступающей расширением к фреймворку TensorFlow Lite, предназначенному для запуска нейросетей на ряде различных устройств помимо ПК. Эта библиотека может быть использована программистом для исполнения моделей на нейроускорителе.

Модель представляется в фреймворке TensorFlow Lite как граф вычислений. В этом графе могут встречаться как и встроенные операторы (по типу `CONV_2D`, `FULLY_CONNECTED` и т.д.), так и пользовательские операторы (“custom operators”), которые программист может создавать и реализовывать в собственном коде [5]. Именно механизмом пользовательских операторов мы воспользовались при разработке данного ПО, ведь он предоставляет наиболее удобный способ интеграции собственного кода в экосистему TensorFlow Lite. Когда код TensorFlow Lite сталкивается с нашим пользовательским оператором, он передаёт контроль коду, производящему действия по координации инференса на нашем устройстве-ускорителе.

Описание ПО для устройства-ускорителя

Программное обеспечение, предназначенное для запуска на устройстве-ускорителе, мы решили базировать на ОСРВ Zephyr, исходя из того, что Zephyr – это активный проект с открытым исходным кодом, предоставляющий встроенные драйвера для многих устройств (в том числе UART-передатчиков) и позволяющий легко менять аппаратное окружение, в котором выполняется код. Программное обеспечение обрабатывает все типы запросов со стороны хоста (`ListInstalledOpsRequest`, `InstallOpRequest`, `InvokeOpRequest`) и отправляет на них ответы согласно протоколу. Как и ПО для устройства-хоста, ПО для устройства-ускорителя написано на языке C++.

Тестирование двустороннего взаимодействия по UART

Мы использовали UART в режиме прерываний, чтобы избежать траты процессорных тактов, возникающей при использовании polling-режима. Из-за

аппаратных ограничений по размеру FIFO, нам пришлось с обеих сторон разбивать передаваемые сообщения на маленькие подпакеты, не превышающие размер FIFO. После передачи каждого подпакета устройство ожидает с другой стороны подтверждающий байт, выступающий сигналом о том, что подпакет принят. Только после получения подтверждающего байта устройство отправляет следующий подпакет. Без данной системы нельзя было бы гарантировать полноту переданных данных, так как вследствие переполнения FIFO некоторые байты могут теряться.

По итогу тестирования программного комплекса, состоящего из всего программного кода, написанного для обеих устройств, мы успешно передали на устройство по UART простую модель, производящую умножение матрицы весов 32x32 на вектор 32x1, и запустили её на центральном процессоре платы, получив на выходе вектор 32x1 с корректными значениями. При этом, как и было задумано, уже загруженная модель не отправлялась повторно.

Заключение

Была успешно реализована первая итерация программно-аппаратного комплекса, нацеленного на обмен данными между отладочной платой и устройством-хостом. Устройства успешно обмениваются данными по интерфейсу USB (с преобразованием в UART). На данном этапе используется интерфейс UART, однако реализованная гибкая объектно-ориентированная архитектура позволяет в будущем воплощать поддержку новых, более быстрых интерфейсов.

Список литературы

1. Official USB 2.0 specification [Electronic resource]. URL: <https://www.usb.org/document-library/usb-20-specification>. (accessed: 01.02.2024)
2. UART specification [Electronic resource]. URL: https://www.infineon.com/dgdl/Infineon-Component_UART_V2.0-Software%20Module%20Datasheets-v02_05-EN.pdf?fileId=8ac78c8c7d0d8da4017d0e7f69251131 (accessed: 02.02.2024)
3. PG142 - AXI UART Lite v2.0 Product Guide (PG142) (v2.0) [Electronic resource]. URL: <https://docs.amd.com/v/u/en-US/pg142-axi-uartlite> (accessed: 03.02.2024)
4. Vivado Design Suite User Guide. Using Constraints [Electronic resource]. URL: https://www.xilinx.com/support/documents/sw_manuals/xilinx2022_1/ug903-vivado-using-constraints.pdf (accessed: 05.02.2024)
5. TensorFlow Lite Guide [Electronic resource]. URL: https://www.tensorflow.org/lite/guide/ops_custom (accessed: 02.03.2024)

АНАЛИЗ АППАРАТНЫХ МЕТОДОВ АВТОМАТИЧЕСКОЙ ОПТИЧЕСКОЙ ИНСПЕКЦИИ ПЕЧАТНЫХ ПЛАТ И УЗЛОВ

Рахимов Р.М.

*Национальный исследовательский университет
«Высшая школа экономики»,
аспирантская школа по техническим наукам
МИЭМ НИУ ВШЭ*

Аннотация

В данной статье проведен анализ аппаратных методов автоматической оптической инспекции (АОИ) печатных плат и узлов. Она состоит из следующих частей, которые включают в себя: введение; виды и типы дефектов на печатных платах и узлах; основные аппаратные элементы АОИ; примеры и сравнения реализаций систем освещения.

Введение

Производство печатных плат является одной из наиболее динамично развивающихся и инновационных отраслей в электронной промышленности [1]. Для удовлетворения высоких требований к серийному производству электронных компонентов необходимо поддерживать стандарты качества продукции. Автоматическая оптическая инспекция является одним из безопасных методов который не разрушает элементы печатных плат и узлов. Этот метод считается надежным и может заменить человеческие ресурсы, которые подвержены усталости и снижению эффективности при выполнении задач контроля.

Полностью автоматизированная система оптического контроля состоит из аппаратной и программной частей. Эта работа сфокусирована в первую очередь на обзоре аппаратной части и реализаций, которые в большинстве своем включают датчики изображения и настройки освещения, отвечающие за получение цифрового изображения.

Данный тезис описывает часть диссертационной работы, связанной с контролепригодностью печатных плат и узлов. Поэтому, с точки зрения научной новизны, в том числе, благодаря анализу аппаратных методов автоматической оптической инспекции, планируется разработать метод автоматического, топологического проектирования, который, учитывает требования по контролепригодности изделия.

Виды и типы дефектов на печатных платах и узлах

При производстве печатных узлов могут возникать различные дефекты, которые влияют на их функциональность. Некоторые из наиболее распространенных дефектов, которые могут быть обнаружены с помощью АОИ, включают:

- дефекты паяных соединений (рис.1): может включать отсутствие паяного соединения, недостаточное количество припоя, избыток припоя, смещение припоя и мостовые дефекты;
- дефекты размещения компонентов: может включать отсутствие, неправильное или двойное размещение компонентов;

- дефекты паяных соединений поверхностного монтажа: может включать дефекты в упаковке микросхем (псевдосоединение, избыток припоя, недостаточное количество припоя, смещение припоя и мостовые дефекты);

- дефекты паяных соединений сквозного монтажа: может включать паяные соединения технологии сквозных отверстий, которые имеют более крупные компоненты и припаиваются с другой стороны печатной платы;

- дефекты золотых пальцев: Золотые пальцы — это соединительные терминалы на краях РСВ, которые могут быть подвержены различным дефектам, таким как сколы, пятна олова, обнаженная никелевая поверхность и необработанные дефекты;

- дефекты печатных дорожек: Печатные дорожки — это медные провода на печатной плате, которые передают электрический ток. Дефекты в этой области могут включать обрыв, короткое замыкание, неправильный размер отверстия, обрыв цепи, слишком близкое расположение проводников, недостаточное травление, избыточное короткое замыкание, отсутствие проводника, отсутствие отверстия, шпиль и чрезмерное травление;

- дефекты BGA (Ball grid array — массив шариков): BGA — это тип поверхностного монтажа, используемый для монтажа электронных компонентов, таких как микросхемы и микропроцессоры. Дефекты, которые могут возникнуть во время процесса производства и размещения BGA, включают отсутствие паяного соединения, недостаточное количество припоя, избыток припоя, смещение припоя и мостовые дефекты.

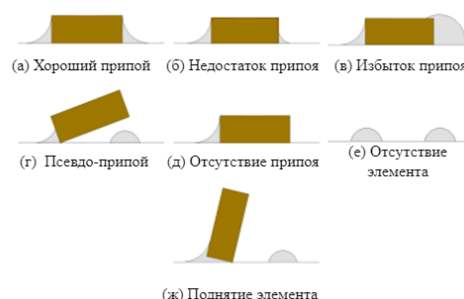


Рис.1. Виды паяных соединений, определяемых АОИ[1]

Основные аппаратные элементы автоматической оптической инспекции

Далее будут рассмотрены примеры реализаций аппаратных решений АОИ (рис.2):

Основные элементы:

1. **Датчик изображения:** Датчик изображения является ключевым элементом системы АОИ [1]. Он отвечает за получение цифрового изображения печатной платы или электронного компонента. Существуют различные типы датчиков изображения, включая ПЗС (прибор с зарядовой связью) и КМОП (комплементарная структура металл-оксид-полупроводник) датчики. ПЗС датчики обычно обеспечивают более высокое качество изображения, но они также более дорогие и требуют больше энергии.

КМОП датчики, с другой стороны, более энергоэффективны и дешевле, но могут иметь более низкое качество изображения.

2. **Настройки освещения:** Настройки освещения также играют важную роль в получении качественного изображения. Они могут включать в себя различные типы освещения, такие как белый свет, ультрафиолетовый свет или инфракрасный свет. Выбор типа освещения зависит от типа дефектов, которые необходимо обнаружить. Например, ультрафиолетовый свет может быть использован для обнаружения дефектов паяных соединений, а инфракрасный свет может быть использован для обнаружения дефектов в светодиодах.

3. **Автоматизация:** Системы АОИ могут быть полностью автоматизированы, что позволяет им работать без вмешательства человека [2]. Может включать в себя автоматическое перемещение печатной платы или электронного компонента, автоматическое изменение настроек освещения, автоматическое обнаружение и классификацию дефектов.

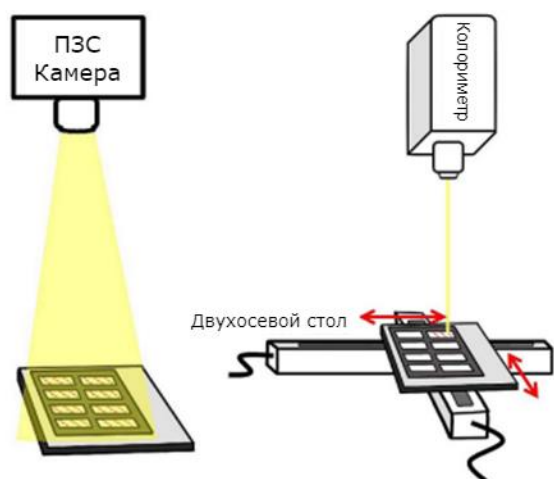


Рис.2. Пример аппаратной реализации АОИ[1]

Примеры и сравнения реализаций систем освещения

• **Трехслойный многоуровневый свет (three-layer tiered light)[1]:** Описание: Данная установка используется для контроля паяных соединений в печатных платах. Установка состоит из 3 (красной, зеленой и синей) круглых ламп в форме кольца на 3 уровнях, построенных по полусферической схеме с различными углами отражения света, которые могут быть зафиксированы или управляться с помощью компьютера. Преимущества: трехмерный характер проверяемого компонента преобразуется в двумерное изображение. Каждый вид дефекта связан с уникальным расположением разноцветных изображений и интенсивностей, что позволяет легко классифицировать дефект. Подходит для проверки паяных соединений. Ограничения: Конфигурация системы очень сложна [3]. Обработка цветных изображений требует большего времени вычислений и сложной обработки. Для получения точных результатов необходима

точная геометрическая калибровка положения источника света.

• **Однослойный многоуровневый свет (single-layer tiered light)[1]:** Описание: Эта установка используется для контроля паяных соединений в печатных платах. Она состоит из многоуровневых светодиодных ламп на с одним источником, а не трех, как в предыдущей установке. Преимущества: Те же, что и в предыдущем варианте, а также низкое время вычислений и более удобная конфигурация системы по сравнению с многослойной настройкой. Ограничения: для достижения хороших результатов необходимо тщательно выбирать и поддерживать угол ϕ [3].

• **Двойной свет (Two lights)[1]:** Описание: В установке использовались два различных источника света: прямой и рассеянный. При прямом освещении источник света поступает непосредственно с относительно большой высоты, а при диффузном освещении источник света равномерно рассеивается перед направлением на паяный шов. Преимущества: Время, необходимое для обработки изображений, невелико по сравнению с многоуровневыми источниками освещения. Ограничения: Система подвержена ошибкам при изменении условий освещения.

Заключение

В статье были описаны и рассмотрены некоторые аппаратные методы автоматической оптической инспекции.

Список литературы

1. Abd Al Rahman, M. A review and analysis of automatic optical inspection and quality monitoring methods in electronics industry / M. Abd Al Rahman, A. Mousavi. – DOI: 10.1109/ACCESS.2020.3029127 – Текст : электронный // Ieee Access. – 2020. – Т. 8 – С. 183192-183271 – URL: <https://ieeexplore.ieee.org/abstract/document/9214824> (дата обращения: 10.02.2024).
2. Дамико, Б. Методы двухмерного и трехмерного контроля на одной платформе автоматической оптической инспекции / Б. Дамико, А. Кожев – Текст : электронный // Технологии в электронной промышленности. – 2013. – № 1(61). – С. 52-54. – <https://elibrary.ru/item.asp?id=18870110> (дата обращения: 10.02.2024). – Режим доступа: Научная электронная библиотека eLIBRARY.RU.
3. Acciani, G. Application of neural networks in optical inspection and classification of solder joints in surface mount technology / G. Acciani, G. Brunetti, G. Fornarelli. – DOI: 10.1109/TII.2006.877265 – Текст : электронный // IEEE Transactions on industrial informatics. – 2006. – Т. 2. – №. 3. – С. 200–209. – URL: <https://ieeexplore.ieee.org/abstract/document/1668079> (дата обращения: 11.02.2024).

ИССЛЕДОВАНИЕ МЕХАНИЧЕСКИХ И ТЕПЛОВЫХ РЕЖИМОВ ЭЛЕКТРОННЫХ ПРИБОРОВ УПРАВЛЕНИЯ ЭЛЕКТРОМЕХАНИЧЕСКИМ ПРИВОДОМ РУЛЕВОГО ДВИГАТЕЛЯ КОСМИЧЕСКОГО КОРАБЛЯ

Шарапов А.Р., Нечухина А.А.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Данная работа посвящена анализу механических и тепловых режимов электронных приборов управления электромеханическим приводом рулевого двигателя космического корабля. Ожидаемым результатом является получение коэффициентов механической и тепловой нагрузки на компоненты данного электронного блока, которые не должны превышать предельно допустимых значений, что способствует повышению надежности и безопасности миссий в космосе.

Введение

В условиях активного развития космической технологии надежное функционирование электромеханических систем управления космических кораблей становится критически важным [1, 2]. Исследование механических и тепловых режимов электронных приборов управления рулевым приводом космических кораблей имеет стратегическое значение для обеспечения безопасности миссий и эффективности работы космических аппаратов (КА). Ведь даже небольшие просчеты или перегрев электронных приборов могут привести к серьезным последствиям, таким как значительное уменьшение времени работы КА или даже полная потеря контроля над кораблем [3].

Целесообразно на ранних этапах проектирования учесть параметры реальных условий эксплуатации и обеспечить пониженные нагрузки электронных компонентов при моделировании объекта, что позволяет снизить риск преждевременных отказов.

Цель работы: исследование устойчивости электронного блока управления электромеханическим приводом рулевого двигателя космического корабля к воздействию внешних тепловых и механических факторов.

В первом разделе работы приводится анализ 3D-модели электронного блока.

Во втором разделе приведены примеры механического и теплового моделирования созданных 3D-моделей, что позволит проанализировать их поведение и характеристики при воздействии внешних факторов.

Создание 3D-моделей электронных приборов управления приводом

На первом этапе работы будет проведено построение 3D-модели электронного блока с использовани-

ем программы САПР Solidworks [4]. Эта программа была выбрана из-за её более широкого распространения в международной практике, что обеспечивает лучшую совместимость с другими программными продуктами и стандартами, а также наличие обширной базы учебных материалов и поддержки сообщества пользователей. Кроме того, Solidworks позволяет достичь высокой точности моделирования, что особенно важно при проектировании таких компонентов, как корпус и платы электронного блока.

Таким образом, в ходе этого этапа была создана детальная 3D-модель всего блока, включая электронные компоненты и приборы управления (рис.1, 2), которые послужат основой для дальнейшего анализа механических свойств и поведения объектов при различных условиях эксплуатации.



Рис.1. 3D-модель электронного блока

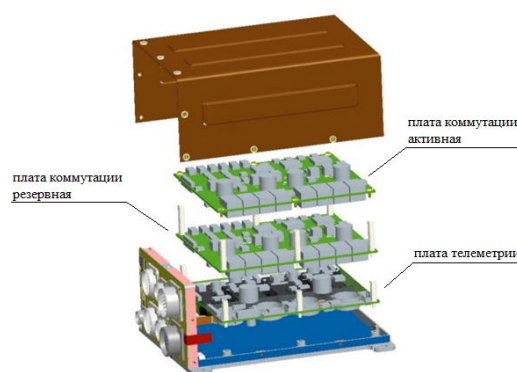


Рис.2. 3D-модель электронного блока с печатными платами

Силовой блок электроники (далее СБЭ) включает в себя три платы. Плата телеметрии собирает данные о работе компонентов двигательной установки в режиме реального времени, она позволяет контролировать параметры работы, такие как давление, температура и т.д. Таким образом, наземные службы получают актуальную информацию о состоянии системы. Плата коммутации активная отвечает за переключение сигналов между различными устройствами или компонентами системы. Она обеспечивает управление и координацию процессов передачи данных и сигналов. Резервная плата коммутации служит на случай отказа основной платы. Она автоматически включается и обеспечивает те же функции, что и активная плата.

Механическое и тепловое моделирование и анализ электронных управляющих приборов

После построения 3D-модели СБЭ в SolidWorks и анализа влияния внешних факторов через крепления, было проведено механическое и тепловое моделирование. Кроме того, после анализа влияния внешних факторов было решено импортировать данные в программу АСНИКА, так как она более удобна для работы с электронными элементами. Для теплового моделирования были выбраны подсистемы АСНИКА-Т и -ТМ, включающие в себя обширный набор инструментов для анализа.

Первым этапом строилась тепловая модель блока и получались температуры конструктивных элементов для задания их в качестве граничных условий для моделирования каждого печатного узла по отдельности. В результате теплового моделирования печатных узлов были получены температуры на каждом электронном компоненте, которые сравнивались с предельно допустимыми значениями температуры из технических условий на них.

Для анализа механических режимов также вначале анализировалась вся конструкция в Solidworks. Полученные значения воздействий на печатные платы через узлы крепления их в блоке необходимы для исследования значений механических ускорений на электронных компонентах и напряжений в материалах уже для печатных узлов в АСНИКА-ТМ.

Исходя из результатов моделирования, производится анализ нагрузок, выдерживаемых электронными компонентами в различных режимах работы КА [6]. На рис.3 и 4 представлены результаты моделирования тепловых и механических режимов, сформированных в подсистеме АСНИКА-ТМ. В приложенных картах представляет интерес столбец коэффициентов тепловой и механической нагрузок. Данные коэффициенты не должны превышать значение 0,7, так как блок разрабатывается для применения в условиях космоса.

На основе полученных данных о нагрузках и напряжениях в материалах платы и электронных компонентах, проводится дальнейший анализ для определения необходимых усовершенствований и модификаций конструкции, направленных на повышение стабильности и надежности работы электронных приборов в условиях космических миссий, например утолщение корпуса, внедрение ребер жесткости или новых креплений.

Таким образом, проведенное двухэтапное моделирование позволяет оценить воздействие теплового фактора и различных механических факторов, таких как вибрации, удары и термические изменения на работу электронных приборов [4]. Определение коэффициентов нагрузок имеет важное значение для обеспечения надежности и долговечности электронных компонентов в условиях космического пространства.

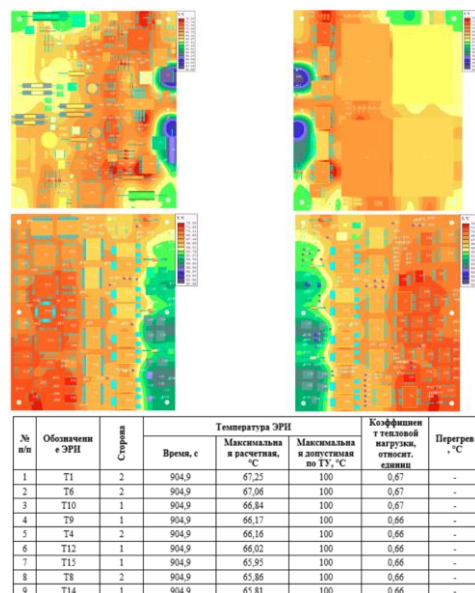


Рис.3. Результаты теплового моделирования. Тепловые поля платы телеметрии, платы коммутации и карта тепловых режимов

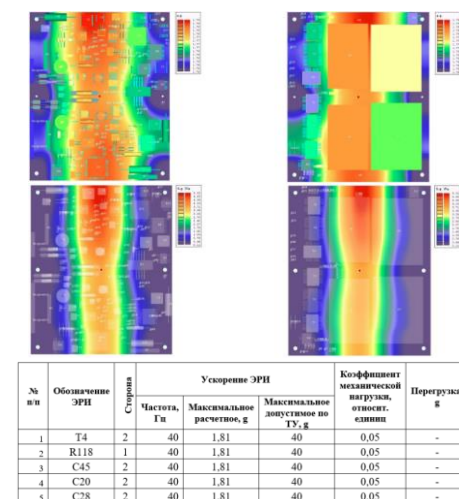


Рис.4. Результаты механического моделирования. Поля ускорений платы телеметрии при воздействии гармонической и случайной вибрации и карта механических режимов

Заключение

Исследование механических и тепловых режимов электронных приборов управления электромеханическим приводом рулевого двигателя космического корабля представляет собой актуальную и перспективную тему для научных исследований. Обеспечение надежности и безопасности работы электронных систем в условиях космического пространства является критически важной задачей, особенно с учетом увеличения числа космических миссий и использования беспилотных аппаратов.

Таким образом, в ходе выполнения данного исследования были обеспечены требования к тепловым и механическим нагрузкам электромеханического привода рулевого двигателя космического корабля

путем проведения компьютерного моделирования в САПР АСОНИКА. Для этого были построены модели печатных плат, а также проведено термическое и механическое исследования аппаратуры с учетом задания параметров внешних воздействий на объект из технической документации. На основе полученных данных были рассчитаны коэффициенты тепловой и механической нагрузок для каждого элемента. Также были применены различные способы по снижению нагрузки на те ЭРИ, у которых значение коэффициента нагрузки превышало 0,7. После применения этих методов были проведены повторные моделирования с учетом внесенных изменений в конструкции, подтверждающие эффективность примененных способов.

Список литературы

1. Грибова Н. В. Основы теории надежности: учебное пособие — Санкт-Петербург : ПГУПС, 2022. 157 с.
2. Big A. A. A. Альмубарак, "Воздействие тепла на электронные компоненты", Международный журнал инженерных исследований и приложений, июнь 2017 г., стр.52-57.
3. А. С. Пиллай, "Введение в ракетную науку и космическое исследование". Бока Ратон, 2022., с. 107-123.
4. Planchard D. S., Planchard M. P. Textbook on SolidWorks 2021: A Step-by-Step Design Approach, 2021.
5. О. В. Широков, С. А. Матвеев, Н. С. Слободзян, А. В. Горбунов "Оценка методов снижения виброактивности в электронных насосных установках космических аппаратов", 2023, стр. 107-114.
6. Кофанов Ю. Н., Полуйко Е. Ю. и Скачко М. А. Метод устранения вибрационных перегрузок в оборудовании космических аппаратов с использованием компьютерного моделирования, в: Инновации, информационные и коммуникационные технологии: сборник трудов XVI Международной научно-практической конференции / Под ред. И. А. Иванова. Ассоциация выпускников и сотрудников ВИА, имени профессора Жуковского, 2019 г., с. 373-378.

РАЗРАБОТКА МАКРОМОДЕЛЕЙ АНАЛОГОВЫХ ИНТЕГРАЛЬНЫХ МИКРОСХЕМ ДЛЯ СХЕМОТЕХНИЧЕСКОГО МОДЕЛИРОВАНИЯ АППАРАТУРЫ

Козаев З.З., Чуча Г.С.

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В рамках работы были измерены статические и динамические характеристики операционного усилителя и счетчика отечественного производства. На их

основе созданы макромодели и проведена их верификация.

Введение

В настоящее время схемотехническое моделирование электронной аппаратуры имеет огромную роль в инженерной деятельности. Оно позволяет не только снизить время проектирования, но также позволяет существенно снизить расходы на проектирование [1]. Моделирование позволяет обеспечить точные и адекватные результаты, а также проводить оптимизацию электрической схемы для повышения эффективности и качества аппаратуры.

Проблема при компьютерном моделировании электронных схем, собранных на основе отечественных компонентов, заключается в том, что в настоящее время не разрабатываются модели отечественных электронных компонентов, что обязывает инженера при моделировании пользоваться иностранными аналогами, что ведет к неточностям в моделировании и увеличению времени проектирования [2]. В рамках данной работы на основе измеренных статических и динамических характеристик были созданы две макромодели электрорадиоизделий, активно используемых в качестве составных элементов электронной аппаратуры: операционного усилителя (ОУ) 544УД16 и счетчика 1564ИЕ19Т.

Создание макромодели операционного усилителя

Статические характеристики измерялись при помощи источника-измерителя Keithley 2612В, а динамические при помощи генератора АКПП-3417, осциллографа АКПП-4134 и источника питания SPS-1230. Так, для ОУ в ходе измерений были получены такие статические параметры, как напряжение смещения нуля $|U_{10}|$, ток питания I_{CC} , средний выходной ток I_1 и разность выходных токов $|I_{10}|$. По измеренным амплитудно-частотным характеристикам (АЧХ) был найден максимальный коэффициент усиления A_U . Все найденные параметры соответствовали спецификации данного электрорадиоизделия [3].

Далее в программе компьютерного моделирования LTspice была создана макромодель на основе модели Бойля [4] (рис.1), значения элементов которой подгонялись так, чтобы параметры, полученные при моделировании в LTspice, совпали с измеренными показателями.

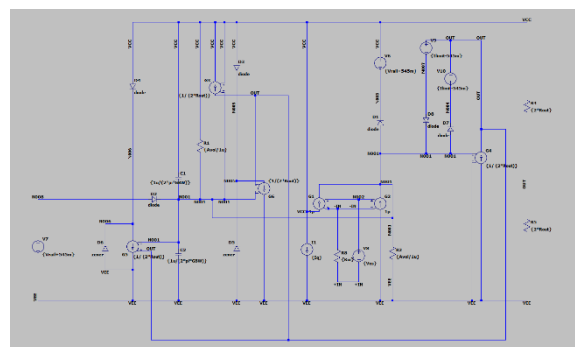


Рис.1. Принципиальная схема модели ОУ 544УД16

Таблица 1. Точность соотношения параметров полученной модели ОУ 544УД16 с измеренными показателями

| | Измерения | Модель | Абс. погр. | Отн. погр, % |
|--|-----------|--------|------------|--------------|
| Макс. коэффициент усиления A_U , дБ | 104,1 | 103,99 | 0,11 | 1,27 |
| Напряжение смещения нуля $ U_{10} $, мВ | 0,1 | 0,112 | 0,012 | 12 |
| Ток питания I_{CC} , мА | 3 | 3,5 | 0,5 | 16 |
| Средний входной ток $ I_I $, нА | 290 | 300 | 10 | 3,4 |
| Разность входных токов $ I_{10} $, нА | 45 | 50 | 5 | 11,1 |

Как можно видеть по таблице 1, относительная погрешность полученных параметров не превышает 16%, что говорит о высокой точности соотношения измеренных и смоделированных параметров ОУ.

Создание макромодели счётчика

Счетчик 1564ИЕ19Т представляет собой 4-х разрядный двоичный счетчик, выходной сигнал которого изменяется по спаду входного. В ходе работы были измерены такие параметры, как ток питания, входной и выходной токи, а также времена нарастания и спада фронтов для всех выходов. Далее была создана макромодель в LTspice (рис.2).

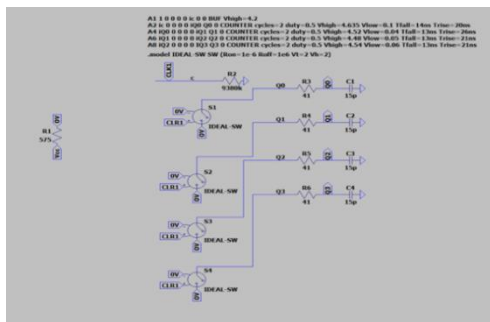


Рис.2. Скриншот макромодели счётчика 1564ИЕ19Т в LTspice

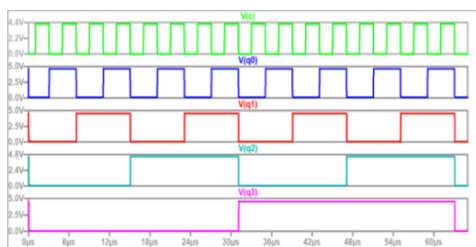


Рис.3. Демонстрация работы макромодели счётчика в LTSpice

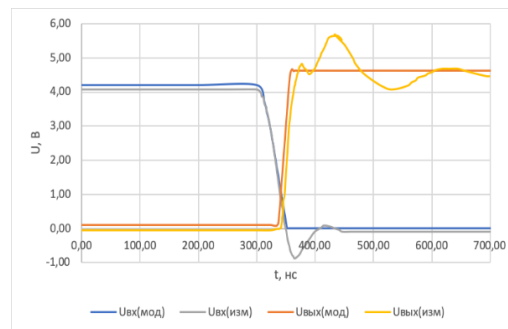


Рис.4. Сравнение измеренного и смоделированного времени нарастания переднего фронта выхода Q0 счётчика

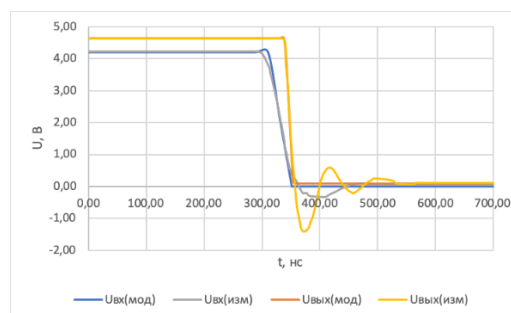


Рис.5. Сравнение измеренного и смоделированного времени спада заднего фронта выхода Q0 счётчика

Таблица 2. Сравнение результатов измерений и моделирования счетчика 1564ИЕ19Т

| Параметр | Измерения | Мо-дель | Абс. погр. | Отн. погр, % |
|------------------|-----------|---------|------------|--------------|
| Ток питания, мА | 7,824 | 7,826 | 0,002 | 0,026 |
| Входной ток, нА | 448,6 | 479,7 | 31,1 | 6,9 |
| Выходной ток, мА | 3,45576 | 3,45636 | 0,0006 | 0,01736 |

По рис.3-5 видно, что измеренные и смоделированные динамические параметры совпадают. Согласно таблице 2, погрешность статических параметров не превышает 7 %, поэтому можно констатировать, что макромодель задана корректно.

Заключение

Были измерены статические и динамические характеристики операционного усилителя 544УД16 и 4-х разрядного двоичного счетчика 1564ИЕ19Т. На основе полученных измерений с помощью программы компьютерного моделирования LTspice были созданы макромодели данных электронных компонентов. Полученные макромодели способны работать в диапазоне температур от -60 до +125 °С.

предназначенная для оценки стоимости жизненного цикла.

Программа оперирует основными индикаторами надежности, среди которых вероятность безотказной работы, средняя наработка на отказ или ремонт и другие. RBD-модуль способен строить несложные структурные схемы, расчет надежности которых основан на методе Монте-Карло. Модуль анализа Парето определяет компоненты и семейство компонентов, которые являются наиболее важными для интенсивности отказов системы и дефектов сборки. Помимо прочего, в ПК входит функциональный модуль построения итоговых отчетов. К отличительным особенностям данной системы также можно отнести наглядное представление информации об изделии, которое достигается средствами графической презентации дерева изделия 1.

ANSYS Sherlock

В рамках исследования надежности электронного изделия, в особенности печатного узла, ПО ANSYS Sherlock использует физические основы надежности, то есть физику отказов на основе метода конечных элементов, причем без построения каких-либо структурных схем. В качестве входных данных выступают файлы EDA, включающие схемы, компоновку и списки деталей, и данные встроенных библиотек.

Программа позволяет выполнить анализ и оценку термических циклов изделия, собственной частоты, гармонической и случайной вибрации, ухудшение рабочих характеристик при действии температуры (см. рис.2), анализ частоты отказов. Результаты работы программы представляются в нескольких видах документированной отчетности: таблицы, гистограммы, а также кривые долговечности, на которых представлен расчет наработки на отказ. Данная программа совместима с множеством CAD-систем, такими как SolidWorks, Autodesk Inventor и другими 2.

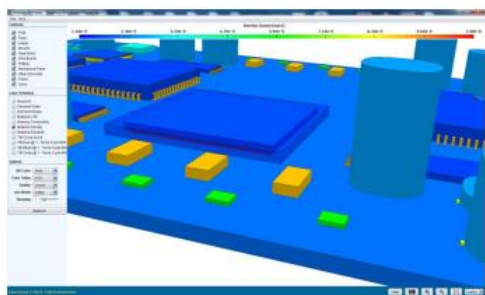


Рис.2. Пример работы ANSYS Sherlock

WQS

ПК WQS (Windchill Quality Solutions), разработанный немецкой компанией KConsult C.I.S., призван для контроля надежности и качества технических систем внутри их жизненного цикла. Этот комплекс встроен в PLM-систему Windchill. Данная программная система получила широкое

распространение для оценки надежности различных объектов промышленной автоматике и потенциально опасных изделий. WQS содержит ряд программных модулей, среди которых можно выделить модуль структурной схемы надежности (RBD; см. рис.3), модуль дерева отказов и событий, модуль марковских процессов, модуль анализа ремонтпригодности и безотказности другие. Последние модули сопряжены с международными стандартами по надежности: MIL-HDBK-217, MIL-HDBK-472, TR-332 и другие. Модуль, связанный с построением и анализом деревьев отказов и событий, имеет особую популярность среди других модулей в связи с анализом развития отказов и событий в системе.

В частности, следует отметить модуль Windchill RBD, позволяющий анализировать надежность и производительность, в том числе рассчитывать основные показатели надежности различных электронных систем, содержащих определенный вид резервирования и восстанавливаемости. В основу расчета индикаторов надежности заложены методы аналитического расчета и методы Монте-Карло. Кроме того, данный модуль поддерживает вложенность блоков схемы и решение задач оптимизации 3.

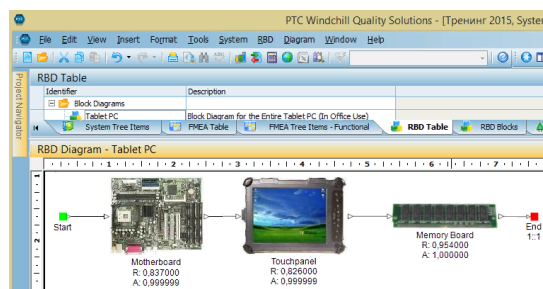


Рис.3. Пример работы WQS

АРБИТР

Первым отечественным программным решением по оценке надежности, который хотелось бы кратко рассмотреть, является ПК АРБИТР, проводящий расчет безопасности и технического риска с помощью методики академика И.А. Рябина, названной общим логико-вероятностным методом (ОЛВМ). Согласно этой методике, строится схема функциональной целостности (СФЦ), представляющая исследуемый объект и позволяющая сформировать аналитические модели надежности и безопасности для расчета основных индикаторов надежности, а также, например, провести анализ риска элементов. Кроме того, данный комплекс имеет модуль по определению минимальных путей успеха и отказа системы, который не был обнаружен в других программных решениях. Его часто используют для оценки надежности больших промышленных объектов, требующих высокой надежности, наподобие атомной электростанции или подводной лодки 4.

АСОНИКА-К

Другим комплексом, который представлен на отечественном рынке, является ПК АСОНИКА-К, схожий по своим особенностям с зарубежными системами RAM Commander и WQS. Ключевым достоинством применения АСОНИКА-К является оценка надежности РЭА, элементы которой имеют отечественное или зарубежное происхождение. Для этого АСОНИКА-К в своей работе опирается на отечественные справочники “Надежность ЭРИ” и “Надежность ЭРИ ИП” (2006 года выпуска), которые соответствуют военному стандарту “Мороз-6”. Для проектирования различной аппаратуры данная программа применяет CALS- и клиент-серверные технологии, что способствует оперативному добавлению новых классов РЭА на сервер. ПК АСОНИКА-К помогает в работе инженерам-схемотехникам и конструкторам в разработке различных радиоэлектронных изделий.

АСОНИКА-К содержит ряд модулей: АСОНИКА-К-СЧ (расчет показателей безотказности и сохраняемости составных частей), АСОНИКА-К-ЗИП (расчет и оптимизация показателей достаточности систем ЗИП), АСОНИКА-К-Д (расчет показателей долговечности), АСОНИКА-К-РЭС (расчет реконфигурируемых изделий) и АСОНИКА-К-СИ (расчет показателей надежности резервированных изделий на основе структурной схемы методом Монте-Карло). Таким образом, данные модули дают минимальную входимость пользователя в программу и соответствие отечественным стандартам по надежности, но (в части АСОНИКА-К-СИ) малую библиотеку групп изделий по виду резервирования и восстанавливаемости 5.

“Надежность”

Отечественный ПК “Надежность”, разработанный компанией “ОКБ Аэрокосмические системы”, используется для снижения стоимости разработки систем путем выявления и исправления потенциально ненадежных компонентов на ранних этапах проектирования, а также оптимизации процессов оценки и повышения надежности, расчета запасных частей и принадлежностей.

Функционально программа делится на несколько модулей: модуль оценки показателей безотказности ЭРИ и материалов (расчет, аналогичный в АСОНИКА-К, по НДВК-217F Notice 2 и справочнику ФГБУ 46 ЦНИИ МО РФ), модуль структурных схем надежности (оформление по ГОСТ Р 51901.14-2007), модуль анализа видов, последствий и критичности отказов, модуль формирования сводки видов и последствий отказов, модуль анализа деревьев неисправностей (по ГОСТ Р 27.302 и Р-4761), модуль анализа общего режима, модуль для расчёта и оптимизации ЗИП, модуль генерации отчетов (по ГОСТ Р 2.105-2019). Результаты первого модуля служат входными данными для других модулей программы 6.

Разрабатываемая программа

На основе полученных данных можно предложить свою программную систему по оценке надежности, включающую некоторые особенности других систем и свои собственные.

1. Мы пришли к выводу, что одним из самых наглядных и эффективных способов расчета основных показателей надежности большинства изделий на основе радиоэлектронных компонентов является построение блок-схемы надежности особенно для резервированных систем.

2. Ввиду большого множества возможных РЭА нельзя выбрать единый метод анализа RBD. Поэтому было предложено объединить возможные виды РЭА по резервированию и восстанавливаемости, а также схеме подключения в виде отдельной библиотеки, основанной на отечественном отраслевом стандарте ОСТ 4Г0.012.242-84, представляющем одно из возможных математических ядер. Другими математическими библиотеками, которые по желанию могут быть выбраны, являются вероятностно-логические и марковские методы.

3. Составление структурных схем будет возможным на удобном графическом редакторе с инструментами добавления/удаления блоков в соответствии с ГОСТ Р МЭК 61078-2021.

4. Одной из ключевых особенностей нашей программы будет минимальная входимость пользователя в теорию надежности.

5. Результатом программы является итоговый отчет в соответствии с требованиями ГОСТ 27.301-95, содержащий все рассчитанные параметры, графики, а также математическую символическую модель, то есть по сути формулу расчета надежности. Такого формата отчета в других программных системах нет.

Заключение

Надежность является одной из самых важных категорий качества объекта, контроль за которой на протяжении всего жизненного цикла позволит повысить конкурентоспособность проектируемого изделия на рынке. Для этого используются наиболее подходящие, в зависимости от свойств изделия и характера расчета, специалисту методы расчета надежности, реализованные в ряде зарубежных и отечественных программных продуктах, обзор которых был проделан в данной работе. Этот обзор помогает сориентироваться на рынке программного обеспечения по оценке надежности.

Можно кратко выделить следующие особенности рассмотренных до этого шести программных продуктов:

1. Данные программные системы могут быть встроены в жизненный цикл изделия и даже быть частью PLM-системы, как например, продукт WQS.

2. Наиболее распространенными решениями по оценке надежности являются построение RBD, деревьев отказов и деревьев событий и анализ методом Монте-Карло.

3. Наиболее отличающимся ПО как внешне, так и по применяемым методам расчета является ANSYS

Sherlock, который основан не на RBD, а на физике отказов с использованием механического и теплового моделирования. Также можно выделить систему АРБИТР, в которой используется оригинальный общий логико-вероятностный метод (ОЛВМ) и использование которой предназначено для сверхсложных объектов техногенного риска.

4. Наименьшую входимость пользователя обеспечивают отечественные программы АСОНИКА-К и “Надежность”, но предлагают довольно ограниченный набор возможных видов изделий в рамках оценки надежности на основе RBD.

По итогу будет создано собственное ПО, в функционал которого входит выбор математического ядра, построение ССН и вывод итогового отчета на основе отечественных стандартов.

Список литературы

1. RAMS Reliability, Availability, Maintainability and Safety Software. - URL: <https://aldservice.com/RAMS-Reliability-Availability-Maintainability-and-Safety-Software.html> (дата обращения: 18.02.2024). - Текст : электронный.

2. Описание WQS. - URL: <http://www.kconsult-cis.com/windchill-quality-solutions.html> (дата обращения: 18.02.2024). - Текст : электронный.

3. Анализ надежности электроники с помощью физики отказов в ANSYS Sherlock. - URL: <http://www.mes-conference.ru/data/year2020/commprnt/Sherlock.pdf> (дата обращения: 18.02.2024). - Текст : электронный.

4. Программный комплекс АРБИТР. - URL: <https://szma.com/wp-content/uploads/2019/02/Arbitr.pdf?ysclid=1f422qgig3579119633> (дата обращения: 18.02.2024). - Текст : электронный.

5. Описание АСОНИКА-К-СИ. URL: <http://asonika-k.ru/products> (дата обращения: 18.02.2024). - Текст : электронный.

6. Руководство оператора ПК “Надежность”. - URL: https://reliability.manufacturing.digital/uploads/manual_operator_reliability.pdf (дата обращения: 18.02.2024). - Текст : электронный

SPICE МОДЕЛИРОВАНИЕ ЭЛЕМЕНТОВ КНС БИС С УЧЕТОМ ТЕХНОЛОГИЧЕСКОГО РАЗБРОСА

Старилова Е.В., Хобов А.П., Исмаил-Заде М.Р.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ

Аннотация

Для отечественного предприятия, специализирующегося на производстве МОП (металл-оксид-полупроводник) КНС (кристалл на сапфире) структур, возникла необходимость создания универсаль-

ных моделей SPICE с учетом технологического разброса. В данной работе представлено построение угловой модели на основе стандартной модели BSIMSOI, а ее возможности проиллюстрированы на примере моделирования ВАХ n -канального и p -канального КНС МОПТ с $W/L = 5,2/1,7$ мкм.

Введение

Одной из важнейших тенденций развития современной КМОП-технологии металл-оксид-полупроводник является уменьшение размеров транзисторов и других элементов интегральных схем («масштабирование»). Более того, даже для хорошо отлаженных, управляемых и устойчивых процессов статистический разброс параметров этих элементов неизбежно увеличивается [1]–[3]. Разброс параметров транзисторов в интегральных схемах является результатом пространственных и временных флуктуаций технологического процесса, таких как колебания легирующих примесей, шероховатость кромок затвора и др.

Разброс технологических параметров приводит к разбросу параметров электронных компонентов, что приводит к разбросу параметров ИС (интегральных схем) в целом [4], [5], и, как следствие, ухудшается соотношение точности, быстродействия и энергопотребления [6].

Целью данного исследования является разработка универсальных SPICE моделей с учетом технологического разброса параметров транзисторов.

Компьютерное моделирование электронных компонентов с учетом разброса параметров элементов дает следующие возможности [7]:

1) Снижение чувствительности схемы к разбросу параметров физической структуры за счет оптимального подбора размеров транзисторов и их взаимного расположения.

2) Повышение параметрической надежности разрабатываемых электронных устройств.

3) Повышение достоверности схемотехнического проектирования.

Экспериментальное исследование ВАХ КНС МОПТ

Были исследованы результаты измерений электрических характеристик КНС МОП-транзисторов с минимальной длиной канала $L_{min} = 1,7$ мкм и напряжением питания 7 В. Тестовая структура включает в себя n - и p -канальные транзисторы с размерами: диапазон ширины каналов $W = 5,2 \dots 100$ мкм; диапазон длин каналов $L = 1,7 \dots 40$ мкм. Каждая тестовая структура содержит 92 кадра.

Использовались усредненные измеренные характеристики для тестовых транзисторов каждого размера. На рис.1 приведен пример таких характеристик для n - и p -канальных КНС МОПТ с $W/L = 100/1,7$ мкм.

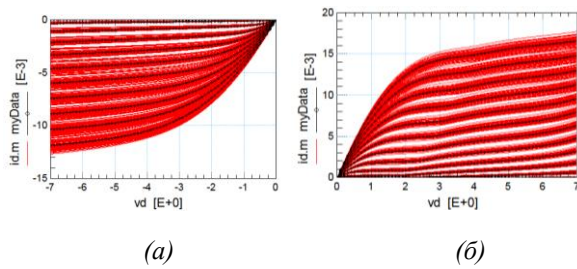


Рис.1. Усреднённые измеренные ВАХ для n-канального (а) и p-канального (б) КНС МОПТ (красные линии – измерения по 92 кадрам, черные линии – среднее значение)

Статистический анализ

Существует статистический локальный и глобальный разброс параметров транзистора, и чтобы учесть его, нужно обратиться к основным параметрам транзисторов, наиболее чувствительных к технологическому разбросу. К таким параметрам относятся: подвижность μ_{eff} , пороговое напряжение V_{th} и ток насыщения I_{dsat} . Чтобы упростить и ускорить статистический анализ, обычно выбираются 3-4 наиболее значимых технологических параметра, остальные считаются фиксированными, что снижает сложность анализа.

Далее необходимые данные получаются путем статистической обработки значений, включая в себя расчет среднего арифметического, СКО, коэффициента вариации, дисперсии и доверительного интервала (рис.2, 4). Коэффициент вариации параметров составляет менее 10%, что указывает на низкий уровень расхождения параметров, а значит позволяет более уверенно прогнозировать поведение интегральных схем.

| Параметр | Подвижность - U0 | | Пороговое напряжение - Vth | | Ток насыщения - Idsat | |
|------------------------|------------------|--------|----------------------------|-------|-----------------------|----------|
| | nfet | pfet | nfet | pfet | nfet | pfet |
| Среднее арифметическое | 463,80 | 300,21 | 1,10 | 1,20 | 1,04E-03 | 7,76E-04 |
| СКО | 20,25 | 12,17 | 0,11 | 0,03 | 3,49E-05 | 2,15E-05 |
| Коэффициент вариации | 4,37% | 4,05% | 10,00% | 2,43% | 3,36% | 2,77% |
| Дисперсия | 410,09 | 148,18 | 0,01 | 0,00 | 1,22E-09 | 4,61E-10 |
| Доверительный интервал | 4,58 | 2,75 | 0,02 | 0,01 | 7,90E-06 | 4,86E-06 |

Рис.2. Основные статистические параметры порогового напряжения, подвижности и тока насыщения для n- и p-МОПТ

Согласно рис.3 около 68% всех значений находятся в пределах одного стандартного отклонения ($\mu \pm \sigma$), около 95% - в пределах двух стандартных отклонений ($\mu \pm 2\sigma$), и около 99.7% - в пределах трех стандартных отклонений ($\mu \pm 3\sigma$). Эти интервалы называют "правилом трёх сигм".

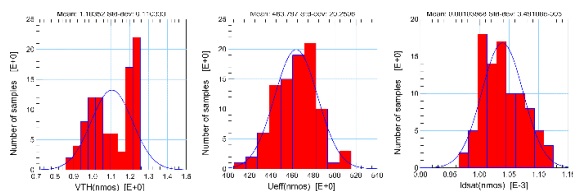


Рис.3. Плотность нормального распределения для порогового напряжения, подвижности и тока насыщения

$$\text{Среднее арифметическое: } \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

$$\text{Стандартное отклонение: } \sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2}$$

$$\text{Коэффициент вариации: } CV = \frac{\sigma}{\bar{X}}$$

$$\text{Дисперсия: } \sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

Доверительный интервал: $\bar{X} \pm t_{\alpha/2} \cdot \frac{\sigma}{\sqrt{N}}$, где $t_{\alpha/2}$ – критическое значение t-распределения для заданного уровня значимости α .

Рис.4. Формулы для расчета статистических параметров

Методика учета технологического разброса в Spice моделях

Первоначально определяются средние значения основных параметров транзистора (μ_{eff} , V_{th} , I_{dsat}), а также рассчитываются средние отклонения σ . Из-за физических и химических процессов во время изготовления транзисторов, возникает технологический разброс, и для того, чтобы учесть его в Spice схемах, строятся угловые модели для данных параметров. В этом методе используется статистический подход для определения крайних точек значений параметров транзисторов. Для аналоговых и цифровых применений создаются четыре угловых случая: SS (медленные nMOS и pMOS) и FF (быстрые nMOS и pMOS) для аналоговых, а также FS (быстрые nMOS и медленные pMOS) и SF (медленные nMOS и быстрые pMOS) для цифровых [8]. Эти точки представляют собой экстремальные случаи, охватывая все возможные промежуточные значения. Для этого параметры смещаются относительно среднего значения (TT) на величину $\pm 3\sigma$, где σ - стандартное отклонение. Такой подход позволяет охватить как наилучший, так и наихудший случай технологического разброса.

Графическое представление угловых точек, построенных на основе экспериментальных данных приведено на рис.5.

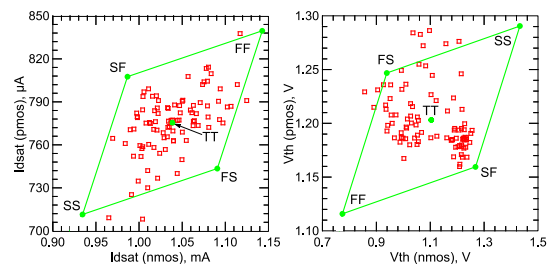


Рис.5. Разброс параметров тока насыщения Idsat и порогового напряжения Vth для n- и p-МОПТ (красные символы – эксперимент, зеленые линии – статистическая модель)

Следующим этапом является определение полного набора параметров Spice моделей для n- и p-МОПТ для случая TT (типичный nMOS и pMOS).

Для учета точек (SS, FF, SF, FS) угловой модели необходимо в базовой SPICE модели TT изменить параметры, которые наиболее чувствительны к технологическому разбросу. Для МОП транзисторов это параметры, отвечающие за пороговое напряжение V_{th} , за ток насыщения I_{dsat} , за подвижность носите-

лей заряда μ_{eff} , за изменение длины L и ширины W канала, за сопротивление сток-исток R_{DS} (см. рис.6).

| Параметр | Значение |
|----------|---|
| Vth0 | Пороговое напряжение |
| U0 | Напряжение переключения |
| Ua | Коэффициент деградации подвижности первого порядка под действием вертикального поля |
| Ub | Коэффициент деградации подвижности второго порядка под действием вертикального поля |
| A0 | Коэффициент зависимости эффекта объемного заряда от длины канала |
| AGS | Коэффициент зависимости объемного заряда от V_{GS} |
| Vsat | Напряжение насыщения |
| RDSW | Сопротивление LDD с нулевым смещением на единицу ширина для RDSMOD=0 |

Рис.6. Параметры модели BSIMSOI

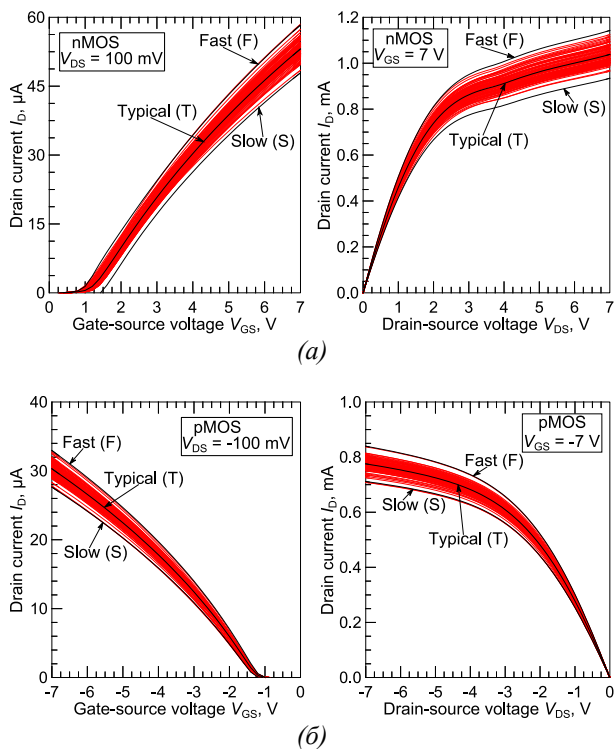


Рис.7. Сопоставление измеренных (красные линии) и смоделированных (черные линии) с помощью угловой модели сток-затворных и выходных ВАХ n -канального (а) и p -канального (б) КНС МОПТ с $W/L = 5,2/1,7$ мкм

Возможности представленной угловой модели на основе стандартной модели BSIMSOIv.4 проиллюстрированы на примере моделирования ВАХ n -канального и p -канального КНС МОПТ с $W/L = 5,2/1,7$ мкм (см. рис.7). Данная модель учитывает большинство эффектов, связанных с уменьшением размеров приборов, таких как эффекты короткого и узкого канала, а также все режимы работы

КНИ (кремний на изоляторе) МОП-транзисторов. BSIMSOIv.4 включает в себя встроенные зависимости от геометрических и технологических параметров МОП-транзисторов, что позволяет охватить весь диапазон длин и ширин каналов в рамках одного набора параметров для всей тестовой структуры.

Заключение

В данной работе была разработана универсальная SPICE-модель для МОП КНС структур отечественного предприятия, учитывающая технологический разброс параметров. В ходе выполнения работы достигнуты следующие результаты. Проведены обработка и анализ измеренных электрических характеристик большого набора тестовых КНС МОПТ по 92 кадрам (рис.1). Выполнен статистический анализ основных параметров транзисторов, наиболее чувствительных к технологическому разбросу (рис.2-4). Построены угловые модели, показывающие разброс параметров тока насыщения (I_{dsat}) и порогового напряжения (V_{th}) для n - и p -МОПТ (рис.5). Определены параметры для построения SPICE-моделей для n - и p -канальных КНС МОПТ (рис. 6). Возможности представленной угловой модели на основе стандартной модели BSIMSOIv.4 продемонстрированы на примере моделирования вольт-амперных характеристик (ВАХ) n -канального и p -канального КНС МОПТ с $W/L = 5,2/1,7$ мкм (рис.7).

Список литературы

1. Pradeep, K., Poiroux, T., Scheer, P., Juge, A., Gouget, G., & Ghibaudo, G. //IEDM'2018, P. 17.1. 1-17.1. 4.
2. Asenov, Asen, and Khairulmizam Samsudin. "Variability in nanoscale UTB SOI devices and its impact on circuits and systems." Nanoscaled SOI Structures and Devices. Springer, Dordrecht, 2007. 259-302.
3. Mroszczyk, P. and Dudek, P., Tunable CMOS delay gate with improved matching properties. IEEE Transactions on Circuits and Systems I: Regular Papers, 61(9) (2014), pp.2586-2595.
4. Saha, S. K. (2014). Compact MOSFET modeling for process variability-aware VLSI circuit design. IEEE access, 2, 104-115.
5. Pradeep, K., Poiroux, T., Scheer, P., Juge, A., Gouget, G., & Ghibaudo, G. (2018) Analysis and modeling of wafer-level process variability in 28 nm FD-SOI using split C-V measurements. Solid-State Electronics, 145, 19-28.
6. Toro-Frias, A., Martin-Lloret, P., Martín-Martínez, J., Castro-Lopez, R., Roca, E., Rodríguez, R., Nafria, M. & Fernández, F. V. (2016). Reliability simulation for analog ICs: Goals, solutions, and challenges. Integration, 55, 341-348.
7. Patil, S.R. and Gautam, D.K., 2017. Statistical modeling of logic gates and flip-flops for high speed CMOS circuits applications. Silicon, 9(3), pp.371-378.
8. S. Saha, "Compact MOSFET Modeling for Process Variability-Aware VLSI Circuit Design" IEEE Access, vol. 2, pp. 104-115, Feb. 2014.

ОТКРЫТАЯ СИСТЕМА ХРАНЕНИЯ И ОБРАБОТКИ НАБОРА ДАННЫХ КОМБИНАЦИОННЫХ СХЕМ

*Пугач В.П.², Чернявских И.И.¹,
Индюченко Н.А.², Гульев А.В.¹*

*Национальный исследовательский университет
«Высшая школа экономики»,*

¹департамент компьютерной инженерии,

*²департамент прикладной математики
МИЭМ НИУ ВШЭ*

Аннотация

В работе представлено программное обеспечение с открытым исходным кодом для генерации, хранения и анализа комбинационных схем. Оптимизированы созданные ранее методы генерации комбинационных схем и сформирован датасет. Генерация комбинационных схем может осуществляться на различных устройствах. В приложении реализована возможность объединения сгенерированных наборов данных в единое хранилище, а также анализа отказоустойчивости комбинационных схем с использованием различных методов их оценки. Предложены новые возможные методы оценки отказоустойчивости комбинационных схем с использованием машинного обучения.

Введение

Комбинационная схема [1] – это цифровая схема, которая, в зависимости от полученных на входах дискретных сигналов, выдает на выходах однозначно определенные логические сигналы.

В любых современных электронных устройствах применяются комбинационные схемы. И, как и любые другие электронные устройства, комбинационные схемы подвержены возможным сбоям и отказам. Неисправность схемы может приводить к неправильной работе всего устройства, что в свою очередь может приводить к серьезным последствиям.

Существующие методы оценки отказоустойчивости имеют ряд недостатков и проблем. Ввиду повсеместности использования цифровых схем, исследование и создание методов для оценки параметров таких схем, является актуальной темой для научных исследований и имеет большое значение для развития современных технологий.

Таким образом, существует необходимость создания открытого программного обеспечения (ПО), которое позволит генерировать большие объемы данных для проведения исследований по предсказанию параметров комбинационных схем, в том числе, с использованием методов машинного обучения.

Целью выполнения работы является разработка единой системы хранения и обработки цифровых схем, которая должна позволить сократить время их анализа, а также исследование методов оценки отказоустойчивости микросхем.

Одной из основных задач, поставленных перед выполнением работы стала доработка программного обеспечения (ПО) для генерации схем. В частности: реализация возможности сохранения сгенерированных схем в дополнительных форматах, необходимых

для дальнейшего их использования при обучении графовых нейронных сетей [2], а также добавление параметра seed для генерации, позволяющего получать однозначные результаты генерации при использовании одних и тех же заданных параметров.

Была реализована система классов, предоставляющих возможность добавлять поддержку новых команд для утилит, предназначенных для работы с комбинационными схемами. Была создана система обработки ошибок, позволяющая корректно реагировать на возникающие при работе с утилитами исключения.

Был реализован новый метод генерации схем, позволяющий создавать схемы с учётом физических ограничений, а также был доработан алгоритм, использующий в своей работе многопоточность.

Переработан класс графа, на котором основывается представление схем в программе. В рамках дальнейшей работы планируется добавление возможности анализа связности сгенерированных схем, а также доработка веб-составляющей и серверной составляющей ПО для работы с новыми реализованными возможностями программы генератора.

Обзор литературы и существующих аналогов

Одним из популярных аналогов системы хранения и обработки датасетов комбинационных схем является ПО HW2VEC, описанное в работе [3]. ПО предоставляет автоматизированный конвейер для извлечения графового представления из аппаратных проектов. HW2VEC включает компоненты для применения графового обучения к этим проектам. Датасет включает в себя Verilog файлы различных аппаратных дизайнов и data-flow графы, сгенерированные для каждой схемы.

Основным недостатком является то, что для получения доступа к данному датасету необходимо приобрести подписку на IEEE DataPort. Сайт предоставляет возможность приобретения дорогостоящей университетской подписки.

Другим аналогом разрабатываемого ПО является продукт Virtuoso Studio, разработанный Cadence [4]. Virtuoso Studio ускоряет проектирование и верификацию интегральных схем и гетерогенных систем. Это решение интегрируется с другими решениями Cadence, предоставляя возможность использования искусственного интеллекта (ИИ) и облачных технологий. То есть программа не объединяет в себе эти функции, а лишь позволяет использовать ПО совместно с другими решениями. Для получения доступа к ПО также необходимо приобрести годовую подписку.

Поиск и анализ существующих приложений и программ для генерации комбинационных схем показал, что в настоящее время не существует полноценных открытых и бесплатных систем, позволяющих генерировать комбинационные схемы при помощи различных алгоритмов генерации, предоставлять доступ к различным представлениям схем, а также оценивать их параметры.

Архитектура программного обеспечения

В данном исследовании было доработано программное обеспечение для генерации комбинацион-

ных схем [5]. Для обеспечения упрощенной процедуры добавления новых методов генерации комбинационных схем, была разработана специальная структура взаимодействия методов генерации с генератором датасета, а также составлена инструкция для добавления новых типов генерации в программу.

Архитектура разработанного клиент-серверного приложения представлена на рис.1.

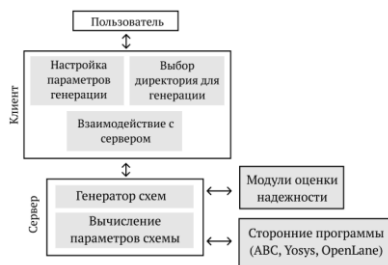


Рис.1. Архитектура разрабатываемого приложения

Программное обеспечение состоит из двух основных составляющих – клиент и сервер, взаимодействующие друг с другом при помощи API. Frontend-часть позволяет осуществлять настройку параметров генерации датасетов и выбор директория для их агрегации, в то время как backend-часть осуществляет генерацию датасетов цифровых схем, а также имеет возможность вычислять параметры схемы, в частности, её отказоустойчивость.

Разработка клиента веб-приложения

Ведется доработка веб-приложения, взаимодействующего при помощи сервера с программой генерации датасетов комбинационных схем и выдающее результаты генерации пользователю на сайте. Пользователь задает параметры генерации, которые сохраняются на сервере. Тем самым он может не переживать за их сохранность. Сама генерация датасетов также происходит асинхронно на сервере, что убирает необходимость для пользователя держать запущенной какую-либо программу или сайт. Используя сервер с генерацией, имеется потенциальная возможность использовать разрабатываемый сервис с различных устройств, на которых каждый пользователь имеет уникальный аккаунт с уникальными данными генерации. В результате все генерируемые схемы агрегируются в облачном хранилище (Synology Drive).

Имеется возможность получения актуальных данных и их обновления во время работы сайта, а также получения статуса готовности генерации датасета.

Разработка серверной составляющей

В работе реализована серверная составляющая, умеющая передавать и принимать данные как веб-приложения, так и программы генерации. Общение с веб-приложением происходит при помощи API.

Сервер разрабатывается на языке Python3, в то время как программа генерации схем написана на языке C++. Модульность программного обеспечения является одним из преимуществ системы.

Общение с программой генерации осуществляется аналогично общению с серверной частью. Несмотря на то, что планируется возможность запускать их на одном компьютере, это не должно стать проблемой, в отличие от использования системных прерываний. Для передачи данных по сокетах на компьютере есть достаточное количество портов, благодаря чему запуск программ на одном устройстве не становится преградой. Также это даёт масштабированность: в дальнейшем планируется осуществить возможность запускать программы генерации, программой предсказания и серверной части, лишь поменяв адрес взаимодействия сокетов.

Доработка программы генератора

Предсказывать отказоустойчивость планируется с помощью изученных алгоритмов [6-11], в частности, алгоритмов с применением машинного обучения. Планируется использовать, как стандартные ML модели с блендингом и стекингом, так и нейронные сети.

Перед реализацией графовых нейронных сетей, необходимо доработать ПО для генерации датасетов, на которых будет производиться обучение моделей.

Была реализована возможность сохранения схем в формате fir rtl [12].

Для тестирования работы алгоритмов генерации, а также для реализации возможности получения однозначных результатов генерации при введении одних и тех же параметров, была добавлена переменная seed. Благодаря добавлению этой переменной пользователь обладает некоторым контролем над генерацией не только по части ввода параметров для генерации, но и на этапе получения сгенерированных схем по введенным параметрам.

Также был реализован новый алгоритм генерации схем, основанный на уже реализованном методе RandLevel, который поддерживает использование логических элементов с заданным количеством входов. Можно сказать, что новый метод генерации позволяет создавать схемы, приближенные по своей сути к тому, как обычно создаются схемы, то есть с учётом физических ограничений.

Для оптимизации параллельной генерации схем был добавлен ThreadPool, позволяющий сократить большое, не оптимальное количество потоков, ограничивая их число и выполняя функции поочередно.

Важной задачей для реализации сохранения сгенерированных схем в необходимых промежуточных форматах, стала переработка классов для графа. В частности, было оптимизировано избыточное копирование, требующее значительных затрат памяти на больших схемах.

Для дальнейших разработок было принято решение существенно обновить классы графа, изменив их структуру так, чтобы каждая из вершин графа могла представлять собой подграф.

Также, за счет добавления словаря, скорость поиска некоторых объектов была уменьшена до $O(1)$, а за счет замены составления имени нового логического элемента из имен входа на хэш, было уменьшено количество требуемой памяти.

Вследствие этих модификаций скорость генерации должна значительно увеличиться, а также это должно позволить полноценно поддерживать поддержку подграфовых модулей для Verilog.

Реализация возможности взаимодействия со сторонними программами

Для добавления возможности взаимодействия со сторонними программами (ABC [13], Yosys [14], OpenLane [15]), была реализована система классов, позволяющая достаточно быстро добавлять поддержку новых команд для утилит.

В частности, с помощью класса, предназначенного для работы с ABC, было реализовано получение данных о задержке распространения в обрабатываемой схеме.

Кроме того, была создана система обработки ошибок, позволяющая корректно реагировать на возникающие при работе с утилитами исключения.

Заключение

Таким образом, результатом данного исследования является доработанное открытое клиент-серверное приложение, позволяющее генерировать комбинационные схемы при помощи различных методов генерации, агрегировать их в выбранном месте, а также обрабатывать датасеты микроэлектронных схем на различных устройствах. В программе предусмотрен функционал для оценки отказоустойчивости микроэлектронных схем и для визуализации сгенерированных схем при помощи другого открытого ПО Yosys.

Разработан CMake файл и организован CI/CD метод разработки для оптимизации работы ПО. Это повысило надежность внесений изменений в программу. На данный момент разработано сохранение в промежуточном формате fig rtl, и ведется доработка классов графа для сохранения в промежуточном формате графа, для дальнейшего использования генерируемых датасетов для обучения графовых нейронных сетей.

Список литературы

1. Харрис Д.М., Харрис С.Л. Цифровая схемотехника и архитектура компьютера: RISC-V / пер. с англ. В. С. Яценкова, А. Ю. Романова; науч. ред. рус. перевода А. Ю. Романов. ДМК Пресс, 2021. с. 199
2. "Графовые нейронные сети," [Электронный ресурс] URL: [education.yandex.ru.https://education.yandex.ru/handbook/ml/article/grafovye-nejronnye-seti](https://education.yandex.ru/handbook/ml/article/grafovye-nejronnye-seti) (дата обращения: 19.02.2024).
3. Rozhin Yasaei, Shih-Yuan Yu, Qingrong Zhou, Mohammad Abdullah Al Faruque, August 26, 2021, "Hardware Design Dataset for Circuit Graph Analysis", IEEE Dataport, doi: <https://dx.doi.org/10.21227/j1vv-hw18>.
4. "Virtuoso Studio," https://www5.cadence.com/virtuoso_studio_info_ip.html (accessed May 31, 2024).
5. CircuitGen [Электронный ресурс] URL: <https://github.com/vvzunin/CircuitGen> (дата обращения: 01.04.2024)
6. Mihir R. Choudhury, Kartik Mohanram, Reliability Analysis of Logic Circuits.// Current approaches to

analysis of the project reliability of electronic devices of cyclic use. // IEEE transactions on computer-aided design of integrated circuits and systems, vol. 28, No. 3, March 2009, p. 392 – 405

7. V.V. Zunin, A.Yu. Romanov, R.A. Solovyev, Developing Methods for Combinational Circuit Generation// 2022 International Russian Automation Conference (RusAutoCon)

8. А.Л. Стемповский, Д.В. Тельпухов, Р.А. Соловьев, Н.В. Тельпухова, Исследование вероятностных методов оценки логической уязвимости комбинационных схем// МЭС-2016. Россия, Москва, октябрь 2016// с. 121 – 126

9. Д.В. Тельпухов, Р.А. Соловьев, Н.В. Тельпухова, А.Н. Щелоков, "Оценка параметра логической чувствительности комбинационной схемы к однократным ошибкам с помощью вероятностных методов"/ Известия Южного федерального университета. Технические науки, No. 7 (180), 2016, с. 149-158.

10. Han J et al. Reliability evaluation of logic circuits using probabilistic gate models. Microelectron Reliab (2010), doi:10.1016/j.microrel.2010.07.154

11. А.Л. Стемповский, Д.В. Тельпухов, Р.А. Соловьев, М.В. Мячиков, Н.В. Тельпухова, Разработка технологически независимых метрик для оценки маскирующих свойств логических схем/ Вычислительные технологии/ Том 21, No 2, 2016, с. 53 – 62

12. Flexible Internal Representation for RTL [Электронный ресурс] URL: <https://github.com/chipsalliance/firrtl> (дата обращения: 19.02.2024)

13. L. Fan and C. Wu, "FPGA technology mapping with adaptive gate decomposition", ACM/SIGDA FPGA International Symposium on FPGAs, 2023.

14. Yosys Open Synthesis Suite, Available at: <https://yosyshq.net/yosys/>, accessed 27.07.2023.

15. OpenLane, [Электронный ресурс] URL: <https://github.com/The-OpenROAD-Project/OpenLane> (дата обращения: 26.01.2023)

АВТОМАТИЗАЦИЯ ПРОЦЕССА МОДЕЛИРОВАНИЯ РАБОТЫ КМОП СХЕМ С УЧЕТОМ ВОЗДЕЙСТВИЯ НА НИХ ВНУТРЕННИХ ФАКТОРОВ СТАРЕНИЯ СО ВРЕМЕНЕМ

Клопотов Г.И.¹, Ивлев К.И.¹, Шундеева А.А.²
Национальный исследовательский университет

«Высшая школа экономики»,

¹департамент электронной инженерии,

²департамент компьютерной инженерии

МИЭМ НИУ ВШЭ

Аннотация

В данной работе проиллюстрирована работа специальной программы для анализа эффектов старения при SPICE моделировании характеристик схемных фрагментов КМОП схем, с проектной нормой 32 нм, при длительной работе схемы.

Введение

Деградация МОП транзисторов вследствие «горячих» носителей при работе СОП схем приводит к постепенному изменению рабочих характеристик электронных компонентов, что значительно снижает время активной эксплуатации компонентов. Постоянный рост требований, предъявляемых к электронным устройствам в настоящее время, обуславливает повышение требований к надежности их компонентной базы [1].

Прогнозирование отказа электронных компонентов при помощи SPICE моделирования схем с учетом факторов старения позволяет подобрать для них подходящие характеристики и режимы работы, что, в свою очередь, позволяет увеличить срок эксплуатации электрической схемы, созданной на их основе.

Новизна настоящей работы обусловлена отсутствием учета эффектов старения в МОП схемах в распространенных пакетах SPICE-анализа.

В работе была поставлена задача расширить возможности программы для автоматизации процесса учета факторов старения в МОП схемах, провести автоматизированный анализ эффектов долговременного изменения параметров МОП транзисторов при работе схем, настроенных на современные проектные нормы, в наибольшей степени влияющих на параметры МОП транзисторов, таких как NBTI, HCI, TDDb и электромиграция, с помощью специально разработанного нами программного обеспечения, получающего из SPICE значения выделяемой мощности компонентов и составляющего тепловые модели компонентов

Эффекты, учитываемые при анализе старения

В процессе анализа материалов из различных источников было выявлено 4 существенных фактора старения МОП транзистора: эффект генерации «горячих» носителей (NBTI) [3], проявляющийся в увеличении порогового напряжения и последующем уменьшении тока стока и проводимости МОП-транзистора, эффект генерации «горячих» электронов и дырок (HCI), эффект пробоя диэлектрика (TDDb), вызванного образованием проводящего пути через оксид затвора к подложке из-за туннельного тока электронов, когда МОП-транзисторы работают вблизи или за пределами их заданных рабочих напряжений, и эффект электромиграции в проводниках ИС, играющий значительную роль в интегральных микросхемах маленького размера.

Создание приложения для расчета параметров МОП транзисторов с учетом постепенных изменений со временем

Использовалось и дорабатывалось программное обеспечение, написанное на языке Python, для автоматизированного вычисления параметров старения компонентов КМОП схем, учитывая влияние рассмотренных факторов старения, с возможностью указания времени работы. В процессе его использования необходимо указать файл нетлиста схемы и выбрать транзисторы, для которых будут проводиться расчеты (см. рис.1). Программа формирует управ-

ляющие команды исходной схемы в LTSpice для расчета параметров старения. Программа считывает результаты моделирования с учетом вставленных команд, обрабатывает их и рассчитывает коэффициенты старения для выбранных компонентов. Далее создается файл с оценкой старения и его графическое представление.

Файл с оценкой старения для схемы на рис.1 представлен на рис.2. Временные диаграммы входного и выходного сигналов и графический расчет эффекта HCI для NMOS представлены на рис.3, 4 и 5 соответственно.

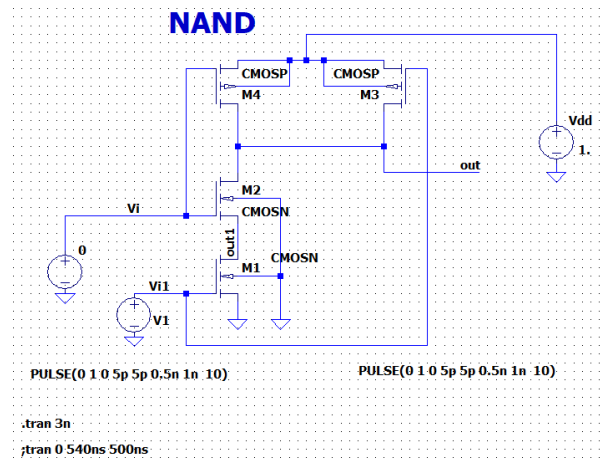


Рис.1. Исходная электрическая схема

```
hci and nbtisRAM_32nm_зарубина 22_01.txt - Блюют
Файл Правка Формат Вид Справка
M6
hci_m6: INTEG(i(b6ib))=-2.75452e-024 FROM 0 TO 1e-008
maxhci_m6: MAX(i(b6ib))=2.09152e-013 FROM 0 TO 1e-008
maxid_m6: MAX(id(m6))=-1.74141e-005 FROM 0 TO 1e-008
M5
hci_m5: INTEG(i(b5ib))=-7.48743e-025 FROM 0 TO 1e-008
maxhci_m5: MAX(i(b5ib))=1.0775e-013 FROM 0 TO 1e-008
maxid_m5: MAX(id(m5))=-6.06324e-005 FROM 0 TO 1e-008

Время работы в режиме NBTI
Время моделирования семы

Отношение времени работы в режиме NBTI к времени моделирования
Сдвиги порога в порядке убывания во время NBTI
[]
Число, на которое надо разделить параметр i0 модели P-канальных МОПТ
Время работы в режиме HCI
M6 1.7559924520603835e-11
M5 -9.265187935034804e-12
Сдвиги порога в порядке убывания во время скачков
['M5 nan', 'M6 0.00263716']
Число, на которое надо разделить параметр i0 модели N-канальных МОПТ
M6 1.0082746798956201
M5 nan
Время моделирования схемы
1e-008
Отношение времени старения к времени моделирования
['M6 0.00175599', 'M5 -0.00092652']
Температура при которой моделируется схема в градусах Цельсия
27
Температура при которой моделируется схема Кельвина
300
Время моделирования схемы
1e-008
```

Рис.2. Результативный файл с расчетом изменения параметров выбранных МОП транзисторов

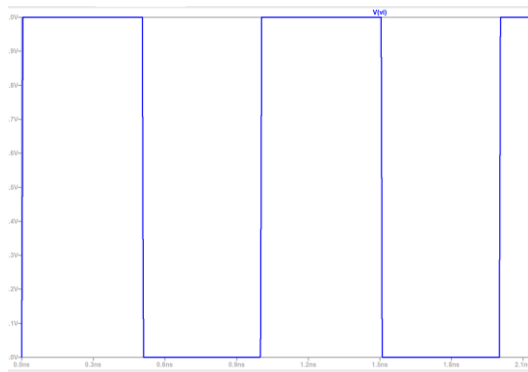


Рис.3. Форма сигнала на входе Vi и Vi1

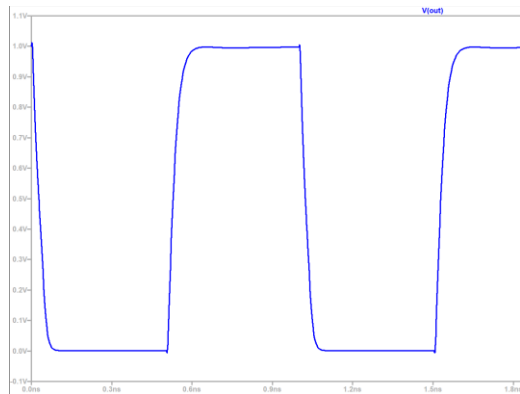


Рис.4. Форма сигнала на Vout

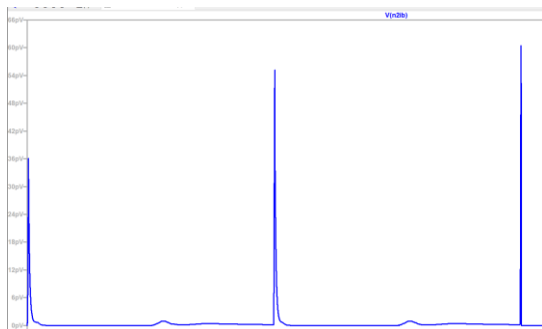


Рис.5. Графическое отображение расчета формул HCl эффекта для NMOS

Формулы для расчета эффектов старения записываются в виде SPICE команд в файл .inc в ходе работы программы. Далее данный файл подключается в LTspice с помощью команды .include, после чего проводятся расчёты. Пример файла для расчета эффектов старения (наглядный процесс расчета эффекта HCl программой LTspice по одной из формул, заданной в файле .include, представлен на рис.5 приведён на рис.6.

```

NAND_12_02_23_1GHz_(1)-1.inc - Блокнот
File Edit Format View Splice
B1Ib 0 M1Ib I=(1e-5)*Id(M1)*pwr((1/32.0), 2.5)*exp(5.7*(V(out1)-V(0)))
R1Ib M1Ib 0 1
.measure tran HCl_M1 integ I(B1Ib)
.measure tran maxHCl_M1 max I(B1Ib)
.measure tran maxId_M1 max Id(M1)
.measure tran maxNBTIM3 max ((V(N001)-V(Vi1))*(V(out)-V(Vi1)))
.measure tran NBTIM3 integ ABS(V(N001)-V(Vi1))
B2Ib 0 M2Ib I=(1e-5)*Id(M2)*pwr((1/32.0), 2.5)*exp(5.7*(V(out)-V(out1)))
R2Ib M2Ib 0 1
.measure tran HCl_M2 integ I(B2Ib)
.measure tran maxHCl_M2 max I(B2Ib)
.measure tran maxId_M2 max Id(M2)
.measure tran maxNBTIM4 max ((V(N001)-V(Vi1))*(V(out)-V(Vi1)))
.measure tran NBTIM4 integ ABS(V(N001)-V(Vi1))

```

Рис.6. Содержимое файла с командами для расчета эффектов старения

Для наглядного представления результатов работы программы были рассчитаны изменения параметров МОП-транзисторов .

NAND 10 ЛЕТ СТАРЕНИЯ

+vth0 = -0.460

M3 -0.460-0.01387344= -0.474

M4 -0.460-0.01387344= -0.474

Число, на которое надо разделить параметр u0 модели P-канальных МОПТ

M3 1.0410435896670207

M4 1.0410435896670207

HCl ДЛЯ NAND

+vth0 = 0.5438

M1 0.04116766+0.5438=0.585

M2 0.04116766+0.5438=0.585

Число, на которое надо разделить параметр u0 модели N-канальных МОПТ

M1 1.1291728832299397

M2 1.1291728832299397

В результате подсчетов были получены параметры u0 и vth0 для n- и p-канальных МОП-транзисторов до учета эффектов старения и после (Таблица 1).

Таблица 1. Значения параметров u0 и Vth0 n- и p-канальных МОП-транзисторов до и после расчетов

| Транзисторы \ Параметры | NMOS (M1, M2) | PMOS (M3, M4) |
|-------------------------|---------------|---------------|
| u0 до расчета | 0.035 | 0.00355 |
| u0 после расчета | 0.031 | 0.00341 |
| Vth0 до расчета | 0.5438 | -0.46 |
| Vth0 после расчета | 0.585 | -0.474 |

На основе полученных результатов, были рассчитаны фронты задержки из 0 в 1 для V(out):

Фронт для V(out) (до старения): T⁰¹ задержки из 0 в 1 = 1.0044 ns - 1.0448 ns = 0.0404ns.

Фронт для V(out) (после старения): T⁰¹ задержки из 0 в 1 = 1.0067 ns - 1.0022 ns = 0.0045ns.

После получения итогового результата программа предлагает произвести перерасчет схемы и создает копию текущего `sig` файла для замены параметров.

Заключение

В заключение, стоит отметить, что в данной работе были автоматизированно рассчитаны характеристики КМОП схем с учетом эффектов старения.

Основными направлениями для дальнейшего развития являются увеличение количества учитываемых факторов старения транзисторов и интеграция созданного программного обеспечения с другими приложениями, используемыми для схемотехнического моделирования помимо LtSpice.

Список литературы

1. Usman Khalid, Antonio Mastrandrea, Mauro Olivieri. Effect of NBTI/PBTI aging and process variations on write failures in MOSFET and FinFET flip-flops. 2015

2. Харитонов, И.А., 2021. Расширение возможностей SPICE-подобных программ за счет учета эффектов старения в МОП схемах, обусловленных эффектами горячих носителей, пробоя диэлектрика и электромиграции. МЭС-2021 (issue DOI: 10.31114/2078-7707-2021-4-73-80), ИППМ РАН Date Views 13.11.2021 www.mes-conference.ru/data/year2021/pdf/D088.pdf#:~:text=Традиционно%20выделяют%20следующие%20эффекты%20в,пМОП%20или%20дырок%20в%20р-МОП.

3. Usman Khalid, Antonio Mastrandrea, Mauro Olivieri. Effect of NBTI/PBTI aging and process variations on write failures in MOSFET and FinFET flip-flops. 2015

ОБЗОР ГЕНЕРАТИВНЫХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ ПРОЕКТИРОВАНИЯ ПЕЧАТНОГО УЗЛА

*Кацнельсон А.И., Кофанова М.А., Новиков К.В.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе рассматривается использование генеративных языковых моделей, решающих задачу Question-Answering, при проектировании печатного узла (ПУ). В рамках исследования проведена оценка работы современных генеративных моделей.

Введение

Целью данного исследования является поиск наиболее подходящей для использования при проектировании печатных узлов нейронной сети.

Нами было выделено три основных области, в которых будут оценено качество генерации. По этим предметным областям был составлен датасет формата CheGeKa[1], содержащий вопросы по подбору

компонентов электрической схемы, ее расчету и проектированию и изготовлению печатного узла.

В ходе работы нами будут проанализированы результаты генерации таких моделей, как (GigaChat[2], Chat-GPT[3]), и измерено качество их генерации с помощью метрики BertScore[4].

Системы Question-Answering

Системы question answering (QA) используются повсеместно для автоматического поиска и предоставления ответов на вопросы, заданные на естественном языке. Они могут быть использованы в широком спектре приложений, включая поисковые системы, виртуальных ассистентов, базы знаний и чат-боты.

Использование вопросно-ответной системы при проектировании ПУ имеет ряд преимуществ. В первую очередь речь идет об автоматизации процесса проектирования: генеративные языковые модели могут значительно упростить и ускорить процесс проектирования печатных плат, позволяя генерировать различные варианты трассировки на основе заданных параметров. Модели могут помочь оптимизировать трассировку печатной платы с учетом различных критериев, таких как минимизация размера, улучшение производительности или снижение энергопотребления, и обнаруживать новые, нестандартные решения для конкретных задач, что может привести к инновациям в области проектирования. Также стоит отметить улучшение точности и предсказуемости процесса проектирования, что может сократить количество ошибок и улучшить качество окончательного продукта.

Системы QA играют важную роль в современном мире, облегчая доступ к информации и повышая удобство использования различных технологий и сервисов.

Задача QA и ее решение

Задача QA в обработке естественного языка заключается в создании системы, способной правильно отвечать на вопросы, заданные на естественном языке.

Для решения этой задачи используются различные методы машинного обучения, однако наибольшую популярность имеют модели глубинного обучения, в частности, трансформеры.

Одной из самых известных и влиятельных больших языковых моделей является GPT (Generative Pre-trained Transformer), созданная компанией OpenAI. GPT построена на технологии трансформер[5] для обработки текста и генерации ответов. Обучение этой модели происходит на огромных корпусах текста без задачи, чтобы затем успешно выполнять различные задачи NLP.

Если говорить об особенностях архитектуры, GPT представляет собой последовательность Decoder блоков, что дает ей преимущество при генерации качественного текста. Подробнее ознакомиться с архитектурой можно на рис. 1.

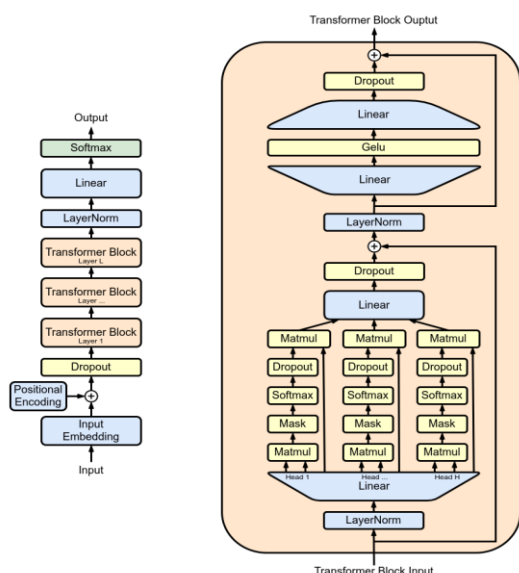


Рис.1. Архитектура GPT

С развитием технологий вышли новые версии модели GPT, такие как GPT-2, GPT-3, ChatGPT и другие. Так, GPT-3 - это одна из самых крупных (из известных) и сложных языковых моделей, содержащая 175 миллиардов параметров. Она способна выполнять широкий спектр задач NLP с высокой точностью и генерировать качественный текст.

Однако чаще всего для улучшения способности модели решать задачу Question Answering используется дообучение модели на специальном образе структурированных наборов данных, таких как SQuAD (Stanford Question Answering Dataset)[6] или MS MARCO (Microsoft Machine Reading COmprehension)[7], которые содержат пары вопрос-ответ для обучения и оценки производительности моделей. На таких же датасетах происходит и оценка модели.

Оценка QA

При оценке качества ответа модели можно использовать разные подходы.

Так, при оценке модели на датасетах с односложными числовыми ответами можно использовать классические метрики, при более сложных вариантах ответов, включающих в себя текст, числа, знаки препинания чаще всего используют метрики, учитывающие семантическую близость слов. Одной из таких метрик является BertScore, которая и была использована в данной работе.

Метрика BertScore основана на использовании предобученных моделей BERT, которые могут извлекать релевантную информацию из контекста и вопроса для предоставления ответа.

Для оценки качества работы модели, сгенерированный и правильный ответы представляются в виде векторов, полученных с помощью модели BERT. Затем метрика BertScore вычисляет сходство между этими векторами, учитывая контекст вопроса и ответа. Метрика состоит из 3 полей: recall, f1-score и precision.

Использование данной метрики позволяет учитывать не только синтаксическое сходство между ответом и эталонным ответом, но и их семантическую близость. Это влечет за собой повышение точности оценки качества ответа системы и помогает выбрать наилучший ответ из множества возможных вариантов.

Тестовые данные

Ввиду того, что не существует готовый датасет только со схемотехническими задачами и вопросами и ответами на них, было принято решение собрать эти данные самостоятельно.

Основой для формата датасета был взят формат CheGeKa. Датасет состоит из следующих информационных полей: instruction (шаблон задания для нейронной сети), text (текст вопроса), topic (тема вопроса) и нескольких полей с мета информацией.

В результате работы были составлены 31 вопрос по теме «подбор компонентов для электрической схемы», 24 вопроса по теме «расчет электрических схем» и 24 вопроса по теме «проектирование и изготовление печатного узла».

Ход работы

Предметом исследования стали следующие модели:

1. ChatGPT-4
2. ChatGPT-3.5 Turbo
3. GigaChat
4. GigaChat Pro

От каждой модели были получены сгенерированные ответы на вопросы, после чего полученные ответы были сравнены с эталонными и на их основе были посчитаны значения метрики BertScore. В таблице 1 представлены усредненные показатели BertScore по моделям.

Таблица 1. Усредненный по моделям BertScore

| | recall | f1-score | precision |
|-------------------|----------|----------|-----------|
| ChatGPT-4 | 0.778224 | 0.750714 | 0.730571 |
| ChatGPT-3.5 Turbo | 0.742704 | 0.708020 | 0.681041 |
| GigaChat | 0.756337 | 0.716449 | 0.684153 |
| GigaChat Pro | 0.752867 | 0.706388 | 0.670153 |

Далее был проведен более детальный анализ качества генерации по темам. Результаты оценки модели ChatGPT-4 представлены в таблице 2.

Таблица 2. BertScore по модели ChatGPT-4 по темам

| | recall | f1-score | precision |
|---|---------|----------|-----------|
| Оценка показателей надежности радио-электронной | 0.83394 | 0.80173 | 0.77505 |

| | | | |
|--|---------|---------|----------|
| аппаратуры | | | |
| Подбор компонентов для электрической схемы | 0.73138 | 0.70529 | 0.69225 |
| Проектирование и изготовление печатного узла | 0.74150 | 0.71429 | 0.69129 |
| Расчёт электрических схем | 0.8313 | 0.80541 | 0.784125 |

Усредненные результаты BertScore по модели ChatGPT-3.5 Turbo по темам можно увидеть в таблице 3.

Таблица 3. BertScore по модели ChatGPT-3.5 Turbo по темам

| | recall | f1-score | precision |
|---|---------|----------|-----------|
| Оценка показателей надежности радиоэлектронной аппаратуры | 0.76515 | 0.73231 | 0.70663 |
| Подбор компонентов для электрической схемы | 0.72264 | 0.68725 | 0.66296 |
| Проектирование и изготовление печатного узла | 0.72333 | 0.69187 | 0.66512 |
| Расчёт электрических схем | 0.77020 | 0.73175 | 0.700042 |

Таблица 4. BertScore по модели GigaChat по темам

| | recall | f1-score | precision |
|---|---------|----------|-----------|
| Оценка показателей надежности радиоэлектронной аппаратуры | 0.79231 | 0.75868 | 0.73094 |
| Подбор компонентов для электрической схемы | 0.73106 | 0.68054 | 0.64125 |
| Проектирование и изготовление печатного узла | 0.71258 | 0.67004 | 0.63458 |
| Расчёт электрических схем | 0.80425 | 0.77579 | 0.752083 |

Таблица 5. BertScore по модели GigaChat Pro по темам

| | recall | f1-score | precision |
|---|---------|----------|-----------|
| Оценка показателей надежности радиоэлектронной аппаратуры | 0.77889 | 0.74405 | 0.71594 |
| Подбор компонентов для электрической схемы | 0.72622 | 0.67454 | 0.63616 |
| Проектирование и изготовление печатного узла | 0.71225 | 0.66945 | 0.63362 |
| Расчёт электрических схем | 0.80729 | 0.75462 | 0.714333 |

Выводы

Результатом исследования является сравнение четырех генеративных моделей от Сбербанка и OpenAI для задач с открытым вопросом в области проектирования печатных. Для оценивания моделей был собран датасет по 4 темам: оценка показателей надежности радиоэлектронной аппаратуры, подбор компонентов для электрической схемы, проектирование и изготовление печатного узла, расчёт электрических схем. После тестирования моделей GigaChat-Lite, GigaChat-Pro, ChatGPT-3.5Turbo и ChatGPT-4. В качестве метрики использовалась нейросетевая метрика BertScore. В результате наилучшие показатели показала модель ChatGPT-4 по всем 4 темам, за ней следует GigaChat-Pro. Примерно одинаковые показатели у ChatGPT-3.5 Turbo и GigaChat-Lite.

Список литературы

1. Taktasheva, Ekaterina & Shavrina, Tatiana & Fenogenova, Alena & Shevelev, Denis & Katricheva, Nadezhda & Tikhonova, Maria & Akhmetgareeva, Albina & Zinkevich, Oleg & Bashmakova, Anastasiia & Iordanskaia, Svetlana & Spiridonova, Alena & Kurenshchikova, Valentina & Artemova, Ekaterina & Mikhailov, Vladislav. (2022). TAPE: Assessing Few-shot Russian Language Understanding. 2472-2497. 10.18653/v1/2022.findings-emnlp.183.
2. developers.sber.ru/gigachat - платформа для создания и развертывания чат-ботов от Сбербанка Электронный ресурс // Sber. (б. д.). URL: <https://developers.sber.ru/gigachat/> (дата обращения: дата обращения).
3. openai.com/chatgpt - сервис ChatGPT от компании OpenAI, предоставляющий возможность использования модели для генерации текста в чатах Электронный ресурс // OpenAI. (б. д.). URL: <https://openai.com/chatgpt> (дата обращения: дата обращения).

4. Zhang, Tianyi & Kishore, Varsha & Wu, Felix & Weinberger, Kilian & Artzi, Yoav. (2019). BERTScore: Evaluating Text Generation with BERT.
5. Dai, Zihang & Yang, Zhilin & Yang, Yiming & Cohen, William & Carbonell, Jaime & Le, Quoc & Salakhutdinov, Ruslan. (2019). Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context.
6. rajpurkar.github.io/SQuAD-explorer - ресурс для исследования и работы с датасетом Stanford Question Answering Dataset (SQuAD) Электронный ресурс // Rajpurkar. (б. д.). URL: <https://rajpurkar.github.io/SQuAD-explorer/> (дата обращения: дата обращения).
7. microsoft.github.io/msmarco - Microsoft Research Open Data Resource for Large Scale Question Answering and Conversational Search Электронный ресурс // Microsoft. (б. д.). URL: <https://microsoft.github.io/msmarco/> (дата обращения: дата обращения).

ОБЗОР ДАТЧИКОВ ПОЛЁТНОГО КОНТРОЛЛЕРА БЕСПИЛОТНОГО ЛЕТАТЕЛЬНОГО АППАРАТА

Дубленский Д.С.¹, Махмудов Т.Н.²
*Национальный исследовательский университет
 «Высшая школа экономики»,
¹департамент компьютерной инженерии,
²департамент электронной инженерии
 МИЭМ НИУ ВШЭ*

Аннотация

В беспилотных летательных аппаратах используются различные типы датчиков, которые используются для систем навигации, стабилизации, сбора данных и др. В зависимости от условий эксплуатации, компоновки, веса, надежности и других параметров существует необходимость в подборе датчиков соответствующим всем предъявляемым требованиям. Данная работа представляет собой обзор доступных решений при подборе датчиков на основании их характеристик.

Введение

В настоящее время разработано множество различных устройств и прототипов для перемещения в воздушном пространстве. Одними из таких механизмов стали беспилотные летательные аппараты. Основным управляющим устройством для таких аппаратов являются полётный контроллер – электронная вычислительная система, которая на основе получаемых с датчиков информации принимает решение по управлению полетом. Цель данной работы – обосновать выбор датчиков, представленных сегодня на рынке, при конструировании полётного контроллера.

Поставленная выше цель позволяет выделить два раздела: теоретические характеристики датчиков и реальная возможность их использования. Таким образом, анализ позволит представить две модели. Первая модель позволит составить план при изуче-

нии внутреннего и теоретического строения датчиков. В то же время вторая модель позволяет взглянуть на существующие требования, предъявляемые со стороны внешнего мира.

В качестве примера применения результатов исследования рассматриваются датчики полётного контроллера беспилотного летательного аппарата, разрабатываемого в МИЭМ НИУ ВШЭ.

Характеристика датчиков

Датчик — это устройство, воспринимающее внешние воздействия и реагирующее на них изменением электрических сигналов [1]. Исходя из такого определения, необходимо выбрать несколько физических величин, на основании значения которых полётный контроллер сможет поддерживать систему в стабильном состоянии. В соответствии с этой идеей, в конструкции БПЛА чаще всего применяются рассмотренные далее датчики.

Каждый датчик необходимо определить. Соответственно, в строящейся модели датчик можно представить в виде формулы. В связи с этим сразу вводятся некоторые обозначения:

- γ – выходной сигнал датчика в вольтах;
- β – систематическая погрешность и влияние температуры;
- η – гауссов шум.

Необходимо уточнить, что у каждого датчика есть определённая систематическая погрешность, выявленная при разработке, а также погрешность, вызванная температурными факторами окружающей среды. Соответственно, общую ошибку этих погрешностей можно представить в качестве их суммы.

Акселерометр – датчик ускорения. В малых летательных аппаратах принято использовать трёхосевой акселерометр. Обычно доступный спектр измерения вычисляется в g (ускорение свободного падения):

$$\gamma_{\text{уск.}} = k_{\text{уск.}}A + \beta_{\text{уск.}} + \eta'_{\text{уск.}}, \quad (1)$$

где $k_{\text{уск.}}$ – это коэффициент усиления, A – ускорение.

Датчики угловой скорости и гироскопы требуются для измерения угловой скорости. Результат измерения представлен градусами в секунду:

$$\gamma_{\text{гироскоп.}} = k_{\text{гироскоп.}}\Omega + \beta_{\text{гироскоп.}} + \eta'_{\text{гироскоп.}}, \quad (2)$$

где $k_{\text{гироскоп.}}$ – коэффициент усиления, преобразующий угловую скорость в вольты, Ω – угловая скорость в рад/с.

Барометр – датчик, измеряющий давление. В рассматриваемой теме барометр потребуются для измерения высоты на основе полученных данных о давлении.

$$\gamma_{\text{давл.}} = \rho gh + \beta_{\text{давл.}} + \eta_{\text{давл.}}, \quad (3)$$

где h – абсолютная высота над поверхностью земли.

Магнитометр – это прибор для измерения напряжённости магнитного поля вдоль трёх перпендикулярных осей, используется для навигации.

$$\gamma_{\text{магн.}} = \psi + \beta_{\text{магн.}} + \eta_{\text{магн.}} \quad (4)$$

где ψ – курсовой угол. [2]

Последним из наиболее популярных датчиков является *GPS (Global Positioning System)* – навигационная система, основанная на спутниках, которая обеспечивает информацию о трёхмерном положении объектов. Приборы данного типа обычно являются подключаемым к полётному контроллеру. Их задача определять местоположение БПЛА или иной точки в пространстве [3]

Стоит отметить, что часто применяется инерциальный измерительный блок, который может включать в себя сразу несколько датчиков. Обычно, в его состав входят гироскоп и акселерометр.

Рассмотренные датчики позволяют получить исчерпывающую информацию о положении БПЛА в пространстве. Другие датчики используются редко и при специфических условиях.

Приведённый краткий анализ позволил выявить несколько важных характеристик:

1. Измеряемый диапазон.
2. Сумма погрешностей.
3. Влияние шума.

Интерфейс – общая граница между двумя функциональными объектами. Изучаемая область делает такое определение более конкретным: интерфейс – общая граница между микроконтроллером и датчиками. Инженерное сообщество предлагает огромное множество различных путей общения между устройствами, поэтому в данном исследовании имеет смысл выделить два основных, которые поддерживаются практически в любом современном датчике: I^2C и SPI . В данном случае важную роль сыграет скорость передачи данных: I^2C предлагает частоту передачи данных до 3,4 МГц, в то время как SPI может достигать больших порядков частоты передачи данных, например, 24 МГц. Также необходимо рассчитать электрические характеристики элементов схемы для выбранного интерфейса с учётом параметров датчиков. Расчёт сопротивления для шины I^2C будет представлен формулами (5) и (6):

$$R_{\text{max}} = \frac{t_r}{0.8473 \times C_b}, \quad (5)$$

$$R_{\text{min}} = \frac{V_{DD} - V_{OL(\text{max})}}{I_{OL}}, \quad (6)$$

где, C_b – суммарная ёмкость выводов устройств, V_{DD} – напряжение питания, t_r – время подъём фронта сигнала V_{OL} – низкий уровень напряжения, I_{OL} – низкий уровень тока. Четыре последних параметра зависят от выбранного режима стандарта I^2C [4].

Ещё одним часто используемым интерфейсом является $UART$. Он применяется при подключении GPS .

Среди прочего, стоит учитывать напряжение питания и требуемую силу тока. Причём датчики могут находиться в разных режимах точности, что влияет на потребляемую энергию.

Внешние факторы

Исследование необходимо дополнить рассмотрением внешних условий, которые напрямую не зависят от внутренней, физической системы отдельного датчика.

Такие требования могут выдвигаться программной средой, которая будет использована для реализации функций разрабатываемого программно-аппаратного комплекса. *Betaflight* – это популярное программное обеспечение, которое позволяет пользователю настроить полётный контроллер, минуя этап программирования [5]. Однако данный программный продукт позволяет подключить магнетометр исключительно по интерфейсу I^2C , согласно рис.1 [6].



Рис. 1. Схема поддерживаемых интерфейсов программным обеспечением *Betaflight* [6]

В дополнение ко всему прочему, нельзя пренебрегать ценой и доступностью при выборе датчика. Возможна ситуация, когда менее эффективный и более дешёвый датчик полностью удовлетворит требованиям поставленной задачи [7].

Таблица 1 представляет конкретные примеры рассмотренных датчиков в ходе работы над полётным контроллером для БПЛА для системы “*OwlEyeUAV*”.

В первую очередь стоит отметить, что программная составляющая проекта связана с архитектурой ПО *Betaflight*, которое, как было замечено выше, позволяет подключить магнетометр только через интерфейс I^2C . И необходимо добавить, что из-за процессов программной составляющей данные, получаемые от акселерометра, барометра, гироскопа и магнитометра, не обрабатываются, пока не получены от всех четырёх датчиков. Из этих двух фактов следует, что максимальная скорость передачи данных по интерфейсу не будет превышать 400 кГц для рассматриваемого случая. Таблица 1 содержит наиболее подходящие по характеристикам датчики. Однако барометр *BMP280* значительно дешевле и выполняет поставленную задачу, *ICM-42688-P* предпочтителен из-за меньших показателей шума и влияния температуры, а *BMM150* имеет более хорошие характеристики в целом. Именно эти датчики пока рассматриваются для проекта “*OwlEyeUAV*”.

Таблица 1. Сравнение датчиков и их характеристик, рассмотренных во время работы над полётным контроллером “OwlEye UAV”

| Тип датчика | Название датчика | Диапазон значений | Погрешность | Влияние температуры | Шум | I^2C | SPI | Ориентировочная стоимость |
|---------------------------------|--------------------|---|----------------|--|--|--------|-------|---------------------------|
| Барометр | <i>BMP280</i> | 300-1100 гПа | ± 1 гПа | 1,5 Па/К | 1,3 Па | + | + | 250 руб. |
| | <i>BMP380</i> | 300-1100 гПа | ± 0.5 гПа | 0,75 Па/К | 0,02 Па | + | + | 2 000 руб. |
| Инерциальный измерительный блок | <i>ICM-42688-P</i> | Акселерометр - $\pm 16g$ Гироскоп - $\pm 2000 dps$ | $\pm 0.5 dps$ | Акселерометр - $\pm 0,15 mg/^{\circ}C$ Гироскоп - $\pm 0,005 dps/^{\circ}C$ | Акселерометр - 0,7 mg Гироскоп - 0,028 dps | + | + | 900 руб. |
| | <i>BMI270</i> | Акселерометр - $\pm 16g$ Гироскоп - $\pm 2000 dps$ | $\pm 5 dps$ | Акселерометр - $\pm 0,25 mg/^{\circ}C$ Гироскоп - $\pm 0,015 dps/^{\circ}C$ | Акселерометр - 1,51 mg Гироскоп - 0,07 $^{\circ} dps$ | + | + | 1 500 руб. |
| Магнитометр | <i>QMC5883L</i> | $\pm 800 \mu T$ | $\pm 1 \mu T$ | $\pm 0,05 \% / K$ | 0,02 μT | + | - | 500 руб. |
| | <i>BMM150</i> | $\pm 1300 \mu T$ | $\pm 40 \mu T$ | $\pm 0,01 \% / K$ | 0,6 μT | + | + | 500 руб. |

Заключение

Беспилотный летательный аппарата – это сложный программно-аппаратный комплекс, который в процессе работы будет полностью полагаться на данные с датчиков. В рамках проведенного обзора подобраны датчики для конкретного БПЛА.

Разработчику следует ориентироваться на следующие пункты при выборе датчиков для полётного контроллера:

1. Изменяемый диапазон величины.
2. Систематическая погрешность измерений датчика, а также систематическая погрешность под влиянием внешней температуры.
3. Влияние Гауссова шума.
4. Интерфейс подключения.
5. Требования по передачи данных ПО к аппаратной составляющей.
6. Цена.

Список литературы

1. Fraden J., King J. G. Handbook of modern sensors: physics, designs, and applications. - New York: springer, 2010. - с. 2.
2. Beard R. W., McLain T. W. Small unmanned aircraft: Theory and practice. - Princeton university press, 2012. - с. 130-155.
3. Tahar K. N., Kamarudin S. S. UAV onboard GPS in positioning determination //The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. - 2016. - Т. 41. - С. 1037-1042.
4. NXP.com: электронная библиотечная система: сайт. - Eindhoven, 2003. - URL: <https://www.nxp.com/docs/en/application-note/AN10216.pdf> (дата обращения: 14.02.2024). - Режим доступа: свободный. - Текст: непосредственный.

5. Betaflight: [Сайт] электронная библиотечная система: сайт. - 2024. - URL: <https://betaflight.com/docs/wiki/getting-started> (дата обращения: 14.02.2024). - Режим доступа: свободный. - Текст: электронный.

6. Betaflight: [Сайт] электронная библиотечная система: сайт. - 2024. - URL: <https://betaflight.com/docs/wiki/getting-started/hardware> (дата обращения: 14.02.2024). - Режим доступа: свободный. - Текст: электронный.

7. Chodorek, A.; Chodorek, R.R.; Sitek, P. Response Time and Intrinsic Information Quality as Criteria for the Selection of Low-Cost Sensors for Use in Mobile Weather Stations //Electronics - 2022. - Т. 11(15).

АНАЛИЗ КИНЕМАТИЧЕСКИХ СИСТЕМ ДЛЯ FDM 3D ПЕЧАТИ: ВЫБОР ЭФФЕКТИВНОГО РЕШЕНИЯ ДЛЯ ПРОИЗВОДСТВА ПРОТОТИПОВ И ФУНКЦИОНАЛЬНЫХ ДЕТАЛЕЙ

Сазыкин Д.Н.¹, Данилов Е.М.¹,
Лебедев М.О.², Клюев Н.С.³

Национальный исследовательский университет
«Высшая школа экономики»,

¹департамент компьютерной инженерии,

²департамент прикладной математики,

³департамент электронной инженерии

МИЭМ НИУ ВШЭ

Аннотация

Работа представляет собой обзор существующих кинематических систем, применяемых в FDM (филаментном наплавлении) 3D печати. В работе рас-

считаются основные типы кинематик, включая дельта, SCARA, CoreXY, H-BOT и параллельные манипуляторы. Каждая кинематическая система описывается с точки зрения принципов работы, технологических преимуществ и перспектив применения. Обзор позволяет оценить сильные и слабые стороны каждой системы, а также определить области их наиболее эффективного использования в различных областях, включая промышленное производство, научные исследования и образование.

Введение

Филаментное наплавление (FDM) является одним из наиболее популярных методов аддитивного производства, который позволяет создавать трехмерные объекты, наращивая материал слой за слоем. Кроме выбора материалов и параметров печати, ключевым аспектом успешной 3D печати является выбор подходящей кинематической системы. Кинематика определяет способ, с помощью которого печатающая головка перемещается по рабочему пространству, что влияет на скорость, точность и качество печатаемых объектов.

Существует множество различных кинематических систем, каждая из которых имеет свои особенности и преимущества. Например, дельта-кинематика обеспечивает высокую скорость и точность печати, в то время как SCARA-кинематика известна своей универсальностью и применимостью в промышленных системах. Кинематика CoreXY и H-BOT также обеспечивает высокую скорость и точность, но использует отличные от остальных принципы перемещения.

Целью данного обзора является систематизация и анализ существующих кинематических систем, применяемых в FDM 3D печати. Будут рассмотрены основные типы кинематик, описаны их принципы работы, технологические преимущества и перспективы применения. Понимание различий между кинематическими системами позволит выбрать наиболее подходящую для конкретных потребностей и задач производства.

Метод FDM

Печать методом послойного наложения (FDM) — это процесс аддитивного производства, реализуемый путем экструзии материалов [1].

Катушка с термопластичной нитью загружается в принтер и при достижении необходимой температуры внутри сопла, из которого подаётся нить, устройство приводится в действие. В процессе филамент проходит через экструдер, расплавляясь и формируя конечное изделие.[2] Также с процессом печати методом FDM можно ознакомиться на рис.1.

Экструдер подключен к 3-осевой системе. Это значит, что она позволяет ему перемещаться по осям X, Y и Z. Уже расплавленный материал выдавливается в виде тонких нитей и соприкасаясь с охлажденной поверхностью предыдущего слоя пластика сплавляется с ним, образуя единую твердую структуру. Для ускорения прогрева обычно расплавленный

пластик охлаждается при помощи дополнительных вентиляторов, которые закреплены на экструдере.

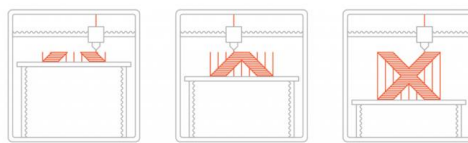


Рис.1. Процесс FDM печати

Картезианская кинематика

Картезианская кинематика (рис.2) является одной из наиболее распространённых и хорошо изученных систем кинематики в контексте FDM 3D печати.

Принцип работы основан на прямоугольной системе координат, где печатающая головка перемещается вдоль трех осей: X, Y и Z. [2]

Основные преимущества, которые можно выделить, это:

- Простота: картезианская система является простой в реализации, эксплуатации и понимании, что делает ее привлекательным выбором для начинающих и любителей 3D печати.
- Надежность: этот тип кинематики обычно характеризуется высокой надежностью и стабильностью работы принтера, что снижает риск образования брака в ходе печати.
- Доступность: многие коммерческие 3D принтеры используют картезианскую кинематику из-за ее относительной дешевизны и простоты в реализации, так как используются универсальные стандартизированные комплектующие

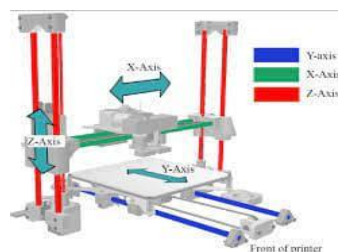


Рис.2. Пример картезианской кинематики

Однако данная система имеет ряд ограничений, которые могут оказаться существенными при выборе кинематики. Во-первых, могут возникать проблемы при печати кривых линий. Во-вторых, скорость печати такой системы может разительно отличаться от аналогов с кинематикой Delta. Это вызвано конструкционной особенностью печатающей конструкции, используемой в обеих системах.

Область применения достаточно обширна. Данный вид кинематики применяется в сфере обучения благодаря своей простоте, доступности и лёгкости освоения. По тем же причинам он используется для прототипирования, а также в серийном производстве, так как принтеры с такой кинематикой являются достаточно эффективными инструментами.

Хотя картезианская кинематика может показаться примитивной по сравнению с некоторыми другими

системами, она все равно является эффективным и широко применяемым вариантом для многих типов 3D печати.

Дельта-кинематика

Дельта-кинематика является уникальным и инновационным подходом к движению печатающей головки в FDM 3D печати.

Принцип работы основан на использовании трёх манипуляторов, расположенных в углах рабочей области, чтобы управлять позицией печатающей головки. Каждый манипулятор состоит из нескольких сочленений, которые работают совместно для перемещения головки в трехмерном пространстве (рис.3). Этот подход обеспечивает высокую степень свободы движения и позволяет печатать на больших высотах. [3]

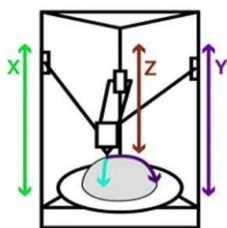


Рис.3. Дельта-кинематика

Основные преимущества, которые можно выделить, это:

- Высокая скорость: дельта-кинематика позволяет достигать высоких скоростей перемещения печатающей головки, что уменьшает время печати (в полтора раза быстрее, чем принтер система с картезианской кинематикой).

- Большая высота печати: благодаря уникальной конструкции, система дельта-кинематики может печатать относительно высокие модели, по сравнению с другими типами кинематики.

И у такой системы есть ограничения. Во-первых, данная конфигурация требует больше времени и навыков для работы с ней. Во-вторых, ограниченная грузоподъемность, что исключает возможность интеграции дополнительных печатающих конструкций. В-третьих, низкая точность относительно других кинематик. Всё это следствие конструктивных особенностей, из-за чего печать возможна только одной печатной головкой и отсутствует возможность обработки более тяжёлых материалов.

Исходя из перечисленных ранее ограничений, данный вид кинематики не так широко используется, однако его можно встретить чаще всего в принтерах направленных на печать высоких объектов или в прототипировании простых наружных форм объектов.

Дельта-кинематика представляет собой инновационный, сложный в конструкции и эффективный тип кинематики, который обеспечивает высокую скорость, точность и область печати в FDM 3D печати. Несмотря на свои ограничения, дельта-принтеры остаются популярным и востребованным инструментом для различных применений от прототипирования до производства деталей.

SCARA-кинематика

SCARA (Selective Compliance Assembly Robot Arm) кинематика представляет собой важный тип конфигурации, широко используемый в промышленных роботах и автоматизированных системах. В последние годы SCARA-принтеры стали получать все больше внимания в мире 3D печати благодаря своей высокой скорости, точности и эффективности.

SCARA-принтеры основаны на принципе работы дельта-роботов и обладают двумя параллельными осями движения (X и Y), а также вертикальным движением по оси Z (рис.4). Это позволяет им эффективно перемещать печатающую головку по плоскости с высокой скоростью и точностью. [4]

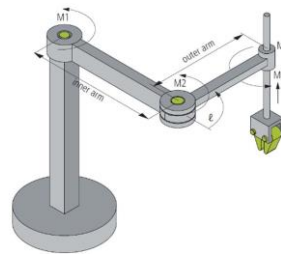


Рис.4. Пример SCARA-кинематики

Основные преимущества, которые можно выделить, это:

- Высокая скорость: В конструкции принтеров данной конфигурации движения печатающей головки происходит по параллельным осям, а управление происходит при помощи эффективной системы управления.

- Высокая точность: благодаря своей конструкции, SCARA-принтеры обеспечивают высокую точность печати, что делает их идеальным выбором для создания деталей с высокой степенью детализации. При сравнении с

- Промышленное применение: SCARA-кинематика широко используется в промышленных системах для роботизированных задач сборки, манипулирования и обработки, что обеспечивает высокий уровень надежности и эффективности.

Основным ограничением этой системы является высота производимых объектов. Оно вызвано в связи с особенностями конструкции принтеров с данной кинематикой.

SCARA-принтеры находят применение в широком спектре областей, включая производство, прототипирование, исследования и образование. Они могут быть использованы для быстрой и точной печати различных объектов, начиная от промышленных деталей до художественных скульптур и моделей.

SCARA-кинематика представляет собой мощный и эффективный способ управления движением в 3D печати. Ее высокая скорость, точность и надежность делают ее привлекательным выбором для широкого спектра промышленных и прототипных приложений.

CoreXY и H-Bot

Кинематики CoreXY и H-Bot (рис.5) являются двумя популярными и инновационными методами

управления движением в FDM (филаментном наплавлении) 3D печати. [5]

Принцип работы основан на использовании двух моторизованных ременных систем для управления движением печатающей головки вдоль осей X и Y. Это позволяет обеспечить высокую скорость и точность печати, за счет минимизации механического напряжения и нагрузки на двигатели.

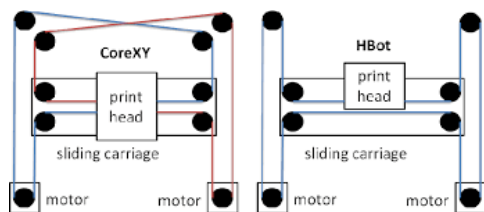


Рис.5. Кинематики CoreXY и H-Bot

Основные преимущества, которые можно выделить, это:

- **Высокая скорость:** CoreXY кинематика обеспечивает высокую скорость перемещения печатающей головки благодаря эффективному использованию моторизованных ременных систем. Одни из самых быстрых 3D FDM принтеров собраны с использованием именно кинематики CoreXY и достигают скорости печати равной около 500 мм/с. В среднем это значение колеблется около 100мм/с и ниже.

- **Высокая точность:** эти кинематики обладают высокой точностью и позволяют создавать детали с необходимой степенью детализации, при подборе правильных параметров в управляющей программе.

- **Жёсткость конструкции и стабильность:** CoreXY и H-bot обеспечивают стабильное и надежное движение печатающей головки, не без помощи жёсткой конструкции, что способствует повышению качества печати. Жесткость конструкции достигается благодаря натяжению ремней, на которых находится печатающая головка.

Единственные ограничения, которые необходимо упомянуть, это сложность монтажа и увеличенное потребление мощности в сравнении с классической картезианской кинематикой. Связано это с использованием двух моторизованных ременных систем.

Кинематики CoreXY и H-Bot находят применение в широком спектре областей, включая производство, прототипирование, робототехнику и образование. Они могут быть эффективно использованы для создания деталей различной сложности и размеров.

Данные кинематики представляют собой эффективные и инновационные методы управления движением в FDM 3D печати. Их высокая скорость, точность и стабильность делают их привлекательным выбором для широкого спектра промышленных и прототипных приложений.

Параллельная кинематика

Параллельные манипуляторы (рис.6) представляют собой тип роботов или манипуляторов, в котором конечные точки (энд-эффекторы) соединены с базой через несколько параллельных кинематических цепей.

Это позволяет им обеспечивать высокую точность, скорость и гибкость в различных промышленных и автоматизированных приложениях.

Кинематика в манипуляторах этого типа основана на использовании нескольких параллельных кинематических цепей, каждая из которых состоит из подвижной платформы, соединенной с базой через серию жестких элементов. Это позволяет энд-эффекторам перемещаться в пространстве с высокой точностью и скоростью, требуемых для выполнения необходимых задач. [6]



Рис.6. Пример Параллельного манипулятора

Основные преимущества, которые можно выделить, это:

- **Высокая точность:** параллельные манипуляторы обеспечивают высокую точность позиционирования и управления, что делает их идеальным выбором для прецизионных задач в промышленности и автоматизации. Однако в точности печати они будут проигрывать принтерам с картезианской системой.

- **Высокая скорость:** благодаря использованию нескольких параллельных кинематических цепей, система может обеспечивать высокую скорость перемещения и обработки объектов.

- **Гибкость и маневренность:** параллельные манипуляторы обладают высокой степенью гибкости и маневренности, что позволяет им эффективно выполнять различные задачи в различных пространственных конфигурациях.

Рассматриваемый вид кинематики имеет достаточно весомые ограничения. Первое и главное – это сложность управления, реализации и настройки. Связано это в первую очередь со сложной и единовременной координацией нескольких кинематических цепей. И второе – это ограниченная грузоподъемность.

Параллельные манипуляторы в первую очередь находят широкое применение в областях робототехники, нежели в аддитивном производстве, однако и в этой сфере на производстве они обширно используются.

Параллельные манипуляторы представляют собой важный класс робототехнических систем, обладающих высокой точностью, скоростью и гибкостью. Однако сфера его использования достаточно узкая из-за специфики конструкции и задач, под которые заточена данная система.

Заключение

В ходе анализа пяти различных типов кинематики в контексте FDM 3D печати, а именно картезианской, дельта, SCARA, CoreXY и параллельного манипулятора, были выявлены их основные принципы работы, преимущества, ограничения и область применения. Каждый из этих методов имеет свои сильные и слабые стороны, а также свой специфический набор преимуществ в зависимости от конкретной задачи.

Кинематика CoreXY выделяется благодаря своей способности обеспечивать высокую скорость, точность и стабильность печати за счет эффективного использования двух моторизованных ременных систем. Этот метод широко применяется в различных областях, включая производство, прототипирование и научные исследования. В результате сравнительного анализа кинематик, CoreXY выделяется как наиболее оптимальный выбор для большинства сценариев 3D печати, благодаря его способности обеспечивать высокую скорость, точность и надежность. Однако выбор кинематики всегда должен зависеть от конкретных потребностей и задач проекта, и в некоторых случаях другие методы кинематики могут быть предпочтительными.

Приведем таблицу, сложенную в течение нашего исследования (таблица 1).

Таблица 1. Сравнительные характеристики различных кинематик

| Тип кинематики | Точность печати | Скорость печати | Качество печати | Сложность конструкции | Надежность конструкции | Сложность калибровки | Возможность модификации | Сложность реализации | Аппаратные ресурсы (штатные драйверы, ИТ) |
|----------------|-----------------|-----------------|-----------------|-----------------------|------------------------|----------------------|-------------------------|----------------------|---|
| Картезианская | Средняя | Средняя | Высокая | Низкая | Средняя | Средняя | Высокая | Средняя | 3 |
| Дельта | Средняя | Средняя | Средняя | Высокая | Низкая | Высокая | Низкая | Высокая | 3 |
| SCARA | Средняя | Средняя | Низкая | Высокая | Низкая | Высокая | Низкая | Высокая | 4 |
| CoreXY | Высокая | Очень высокая | Высокая | Средняя | Высокая | Средняя | Высокая | Средняя | 3+ |
| Параллельная | Очень высокая | Очень низкая | Высокая | Высокая | Очень низкая | Очень высокая | Низкая | Очень высокая | 3+ |

Список литературы

1. Горьков Д. Е. 3D-печать с нуля. – БХВ-Петербург, 2020.
2. Дьяченко Е. А. и др. Анализ существующих кинематик 3D принтеров технологии FFF/FDM //Известия Волгоградского государственного технического университета. – 2021. – №. 1. – С. 57-60.
3. Сапрыкина Н. А., Проскоков А. В., Сапрыкин А. А. Синтез механизмов с кинематикой типа дельта //Фундаментальные и прикладные проблемы техники и технологии. – 2020. – №. 1 (339). – С. 26.
4. Гурский Н. Н., Скудняков Ю. А. Роботизированная система построения 3D объектов. – 2017.
5. Avdeev A. R., Shvets A. A., Torubarov I. S. Investigation of kinematics of 3D printer print head moving systems //Proceedings of the 5th International Conference on Industrial Engineering (ICIE 2019) Volume I 5. – Springer International Publishing, 2020. – С. 461-471.
6. Белоусов А. В. и др. Параллельный FDM-принтер //Математическая физика и компьютерное моделирование. – 2016. – №. 4 (35). – С. 116-131.

ИССЛЕДОВАНИЕ ВОЗДЕЙСТВИЯ ЭЛЕКТРОСТАТИЧЕСКИХ РАЗРЯДОВ НА МОЩНЫЕ МОП ТРАНЗИСТОРЫ С УЧЁТОМ ВЛИЯНИЯ ЁМКОСТИ ПЕЧАТНОЙ ПЛАТЫ

*Кузнецов А.Г., Тумковский С.Р., Пожидяев Е.Д.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В данной работе исследуется влияние воздействия ЭСР на мощные МОП транзисторы серии IRF. Ранее в процессе моделирования было установлено, что воздействие электростатических разрядов оказывает больший эффект на транзисторы, имеющие меньшие ёмкости затвор-исток. Данное исследование проводится с целью экспериментального подтверждения полученных результатов компьютерного моделирования.

Введение

Современная радиоэлектронная аппаратура обладает высокой чувствительностью к электростатическим зарядам, взаимодействие с которыми может привести к отказу компонентов или полному отказу аппаратуры [1][2]. Очевидно, что опасность электростатического разряда (ESD) для полупроводников и интегральных схем (ИС) существует где угодно и когда угодно на протяжении всего срока службы любого продукта микроэлектроники, включая производство, упаковку, транспортировку, установку для полевого применения[3].

В настоящей работе излагается план экспериментального исследования влияния ёмкости печатных плат на отказы, связанные с пробоем подзатворного слоя диэлектрика. Ранее в работе [4] уже было проведено подобное исследование, однако в том случае изучалось влияние ёмкости посадочного места транзистора на его устойчивость к воздействию эср.

В качестве объектов для исследования были выбраны n-канальные транзисторы IRF510, IRF420, IRF520, IRF620, IRF7201, IRF630, имеющие различные ёмкости затвор-исток C_{GS} : 135 пФ [5], 300 пФ [6], 350 пФ [7], 450 пФ [8], 550 пФ [9], 600 пФ [10], соответственно.

Вышеперечисленные транзисторы имеют низкое сопротивление при открытии, высокую скорость переключения и низкое тепловое сопротивление, что позволяет применять их при конструировании в качестве регуляторов мощности, преобразователей тока и напряжения, инверторов, силовых переключателей [11].

Компьютерное моделирование

Моделирование воздействия электростатического разряда было выполнено с использованием системы схемотехнического моделирования *LTspice*, так как она предоставляет необходимый функционал для создания моделей аналоговых и цифровых схем, а также эффективно осуществляет моделирование пе-

реходных процессов [11]. Схемотехническая модель представлена на рис.1.

Характеристики печатной платы, на которой установлен МОП транзистор, определяются значениями индуктивности L_2 и ёмкости C_1 .

Моделирование количественных характеристик импульса электростатического разряда осуществляется путем введения в схему элементов V1, S1, L1, R1 и C2.

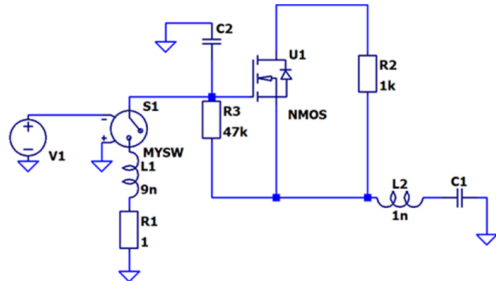


Рис.1. Схемотехническая модель воздействия ЭСР на МОП транзисторы

Экспериментальное исследование будет проводиться при напряжениях разряда U_1 в интервале от 100 до 600 В. Фиксироваться будет переходный процесс изменения напряжения затвор-исток при различных величинах ёмкости ПП.

Кривые изменения напряжения затвор – исток во время переходного процесса при напряжении разряда $V_{esd}=300$ В для транзистора IRF510 показаны на рис.2. Критерием пробоя подзатворного оксида при моделировании является величина 80 В.

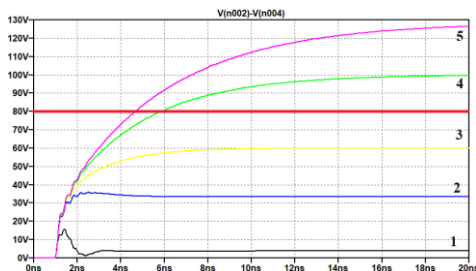


Рис.2. Напряжение затвор-исток во время переходного процесса в транзисторе Irf510 при напряжении разряда $V_{esd} = 300$ В и ёмкостях ПП $C_1 = 5$ пФ (кривая 1), 50 пФ (кривая 2), 100 пФ (кривая 3), 200 пФ (кривая 4), 300 пФ (кривая 5)

В работе ставится задача провести экспериментальные исследования воздействия ЭСР на серию мощных МОП транзисторов с различными ёмкостями затвор-исток при различных значениях ёмкости печатной платы. Диапазон ёмкостей ПП различных размеров и конфигураций был определён из практики.

Экспериментальная часть

Для этого разработан макет экспериментальной установки, представленный на рис.3. Здесь на печатную плату устанавливается испытуемый МОП транзистор, а заданная ёмкость печатной платы определяется ее размерами. С использованием имитатора на затвор транзистора будет подаваться импульс ЭСР

по модели человеческого тела (МЧТ), и фиксироваться переходной процесс напряжения затвор-исток.

Для создания ЭСР будет использован имитатор ЭСР по МЧТ (ГОСТ Р 53734.3.1-2013)[12]. Смоделированные в ПО LTspice формы тока представлены на рис.4.

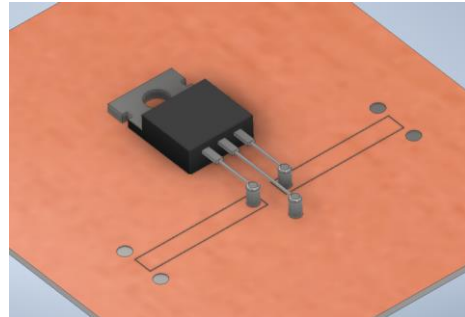


Рис.3. Эскиз макета для проведения эксперимента

Результаты данного исследования в дальнейшем можно будет сопоставить с результатами, полученными путём компьютерного моделирования, что позволит определить степень адекватности использованных spice-моделей в процессе схемотехнического моделирования, оценить погрешности получаемых в результате моделирования данных в сравнении с экспериментальными данными.

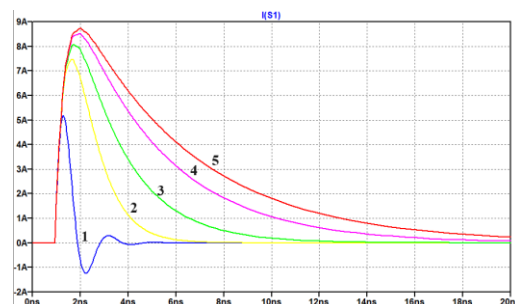


Рис.4. Формы тока при напряжении разряда $V_{ESD} = 300$ В и ёмкостях ПП $C_1 = 5$ пФ (кривая 1), 50 пФ (кривая 2), 100 пФ (кривая 3), 200 пФ (кривая 4), 300 пФ (кривая 5)

Список литературы

1. Кечиев л.н., пожидаев е.д. Защита электронных средств от воздействия статического электричества. М.: ид «технологии», 2004. 346 с.
2. Ke yuan research on esd protection design for electronic communication products. 6th international conference on electronic, mechanical, information and management society (emim 2016). <https://doi.org/10.2991/emim-16.2016.135>
3. Albert wang, "why esd?," in practical esd protection design , ieee, 2022, pp.1-17, <https://doi.org/10.1002/9781119850434.ch1>
4. А.а. Дроздова, и.и. Николаев, м.е. Комнатнов влияние ёмкости посадочного места транзистора на его устойчивость к воздействию электростатического разряда. Журнал радиоэлектроники. Issn 1684-1719. 2023. №1.

5. Irf510 [web]. Hexfet power mosfet. Дата обращения 20.02.2024. Url: <https://pdf1.alldatasheet.com/datasheet-pdf/view/67483/intersil/irf510.html>

6. Irf420 [web]. Hexfet power mosfet. Дата обращения 20.02.2024. Url: <https://pdf1.alldatasheet.com/datasheet-pdf/view/67479/intersil/irf420.html>

7. Irf520 [web]. Hexfet power mosfet. Дата обращения 20.02.2024. Url: <https://pdf1.alldatasheet.com/datasheet-pdf/view/67484/intersil/irf520.html>

8. Irf620 [web]. Hexfet power mosfet. Дата обращения 20.02.2024. Url: <https://pdf1.alldatasheet.com/datasheet-pdf/view/67489/intersil/irf620.html>

9. Irf7201 [web]. Hexfet power mosfet. Дата обращения 20.02.2024. Url: <https://pdf1.alldatasheet.com/datasheet-pdf/view/89534/irf/irf7201.html>

10. Irf630 [web]. Hexfet power mosfet. Дата обращения 20.02.2024. Url: <https://pdf1.alldatasheet.com/datasheet-pdf/view/67490/intersil/irf630.html>

11. Константинов ю.а., горланов е.с., пожидаев е.д., тумковский с.р. Компьютерное моделирование воздействия электростатических разрядов на мощные моп-транзисторы с учетом влияния ёмкости печатной платы. Системный администратор. 2018. №9 (190). С.84-89.

12. Konstantinov u.a., pozhidaev e.d., tumkovskiy s.r. Investigation of electrostatic discharge effect on high-power mosfet-transistors considering the influence of pcb. Int. Seminar on elec. Dev. Design and prod. (sed). Prague, czech republic. 2019. P.1-4. <https://doi.org/10.1109/sed.2019.8798468>

ОБЗОР ОСНОВАННЫХ НА МАТРИЧНОМ УМНОЖЕНИИ АЛГОРИТМОВ СВЁРТКИ НА АППАРАТНОМ УРОВНЕ

Деркач В.В.¹, Батонова О.Ю.², Рожин А.К.³
Национальный исследовательский университет «Высшая школа экономики»,
¹ *департамент компьютерной инженерии,*
² *департамент прикладной математики,*
³ *департамент электронной инженерии*
МИЭМ НИУ ВШЭ

Аннотация

Несмотря на постоянное развитие в области глубокого обучения и появление новых архитектур, сверточные и полносвязные слои остаются основой нейронных сетей [1]. В связи с большим процентом процессорного времени, который приходится на вычисления, связанные с этими слоями [2], ускорение данных операций приводит к значительному увеличению общей производительности. Именно сведение описанных выше аспектов к матричному умножению обеспечивает простоту [2]. В данной работе произведен обзор и сравнение методов, которые способны оптимизировать матричные вычисления, в частности для операций свертки и полносвязных слоев, с аппаратной и алгоритмической точек зрения. Также исследованы модификации алгоритмов, рассчитаны

затраты памяти и необходимые IP ядра для их дальнейшей реализации.

Введение

Рассмотренные алгоритмы используют GeMM (General Matrix Multiply) подход, задача которого сводит свертку к матричному перемножению. Это позволяет распараллеливать вычисления. К тому же, эффективное вычисление больших матрично-векторных и матрично-матричных произведений является хорошо изученной проблемой. BLAS (Basic Linear Algebra Subprograms) содержат подпрограммы для базовых операций с векторами и матрицами, которые находятся в свободном доступе и имеют различные версии [3–5].

Основной идеей алгоритмов является переформатирование данных определенным образом с целью формирования матриц для перемножения. В данной работе проанализированы оба подхода, рассмотрены существующие реализации и модификации, а также произведено сравнение алгоритмов, их недостатков и преимуществ относительно друг друга.

Решение

Image to column (im2col)

Основная концепция состоит в том, чтобы развернуть как входное изображение, так и ядро в вектор, а затем выполнить матричное умножение матрицы веса фильтра W и входной матрицы X .

Для новой матрицы количество столбцов матрицы – это число операций свертки (число окон ядра свертки) для входного изображения, как показано на рис.1.

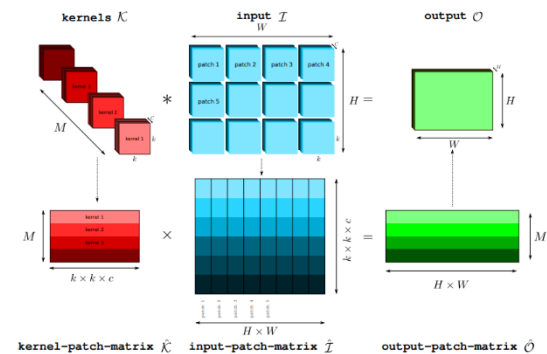


Рис.1. Алгоритм im2col [6]

Алгоритм im2row очень схож с im2col, отличие заключается только в том, по какой размерности раскладывается ядро и изображение. В этом методе окна изображения раскладываются по строкам, а ядра свертки – по столбцам.

Существует несколько возможных вариантов имплементации. Такой метод как наивный подход выполняет явное преобразование im2col для получения уменьшенной карты объектов до того, как она будет использована движком GEMM, в роли которого могут выступать тензорные ядра графических процессоров Nvidia или TPU. Это явное преобразование im2col приводит к значительным затратам произво-

дительности и памяти по сравнению с самим вычислением GEMM [7].

На самом деле ни графические процессоры Nvidia, ни TPU не используют явный `im2col` (хотя эта опция доступна на графических процессорах Nvidia), предположительно, из-за высоких накладных расходов.

Среди прочего рассмотрены следующие алгоритмы имплементации: с использованием кроссбара и на основе буферов. Размер кроссбара определяется как количество строк входного изображения, умноженное на размер патча или вектор ядра. Входное изображение преобразуется для хранения: в каждом столбце будет находиться новая строка изображения со сдвигом на размер ядра. Каждая ячейка памяти кроссбара связана через кроссбар с каждым элементом (выходом) окна изображения. Использование данного метода влечет за собой следующие недостатки: большой размер кроссбара и необходимость преобразования входного изображения для хранения в нужном формате. В свою очередь, в методе, основанном на использовании буферов, создаются $n-1$ линейных буферов, где n – это высота ядра, каждый из которых хранит одну строку исходного изображения и $n \times n$ буферов на один элемент на выходе. Эти буферы последовательно сдвигаются вправо, проталкивая по одному элементу изображения за такт. После некоторого количества установочных тактов, модуль выдает один столбец матрицы патчей за каждый такт, одновременно извлекая один элемент исходного изображения.

В целом подход может сэкономить время доступа к памяти для достижения ускорения, если целью является чтение непрерывной памяти во время вычислений.

Kernel to row (kn2row), kernel to column (kn2col)

При использовании `kn2` алгоритмов в отличие от `im2` алгоритмов не создаются патчи (окна), то есть не копируются данные. На примере алгоритма `kn2row`, данные читаются в определенном формате. Ядро в порядке $K2 \times M \times C$, а изображение - $C \times (H \times W)$. Где C - количество каналов, M - количество ядер (фильтров). В результате получается $K2$ матриц $M \times C$ и одна матрица поступающих данных, где каждая строка - отдельный канал изображения, записанный в виде $(H \times W)$. Затем производится матричное перемножение, и сдвиг перед суммированием по определенному алгоритму, как показано на рис.2.

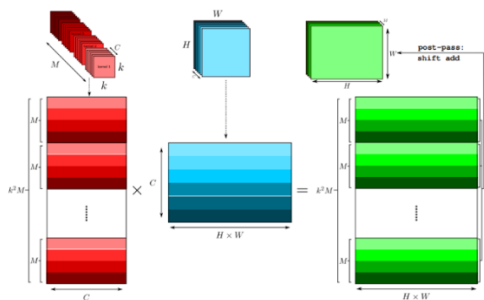


Рис.2. Алгоритм `kn2row` [6]

Таким образом, выходное изображение при нечетном ядре получается такого же размера, что и входное, что равносильно свертке с параметром `SAME`.

Рассмотрим вышеописанный алгоритм на примере матрицы изображения $C=1, H=3, W=3$ и ядра, размерность $k = 3$. На рис.2 (а) показаны данные входного изображения в обычном 2-х мерном представлении, а на рис.2 (б) результат матричного перемножения до окончательного суммирования. Поскольку $C = 1$, то матричное перемножение сводится к перемножению вектора $H \times W$ на каждую отдельную ячейку ядра. Рассмотрим, какие значения исходного изображения участвовали бы при традиционной свертке с параметром `SAME` для каждой ячейки ядра и выделим их соответствующим цветом. Заметим, что центральная ячейка ядра проходит по всем значениям исходного изображения, поэтому удобно начать расчёт смещения с нее. Поскольку при итоговом суммировании в каждом скалярном перемножении должно присутствовать слагаемое с перемножением центральной ячейки и соответствующей ячейки изображения по порядку, то смещение и сдвиг перед суммированием не нужно реализовывать. Перейдем к другим значениям ядра. Рассмотрим верхнюю левую ячейку. Если аналогичным образом выделить область, которую она затронет в свертке, то в результате в первых 4 перемножениях получилось бы нулевое значение, поэтому перед суммированием пропускается 4 ячейки в результирующем векторе.

$$t = x * W + y \quad (1)$$

- t - сдвиг промежуточного результата
- x - насколько взятая ячейка ядра выше центра фильтра (Если ниже, добавляется знак минус)
- y - насколько взятая ячейка ядра левее центра фильтра (Если правее, добавляется знак минус)

Далее необходимо рассчитать пропуски, поскольку при переходе на новую строчку будет осуществляться умножение на 0, а не на последнее значение предыдущей строки. И последняя строка, которая не затрагивается левой верхней ячейкой аннулируется начальным сдвигом, так как количество поэлементных перемножений при прохождении окна ядра по изображению одинаково, то выходит, что сколько первых значений обратилось в нуль, столько элементов в конце не затрагивает определенная ячейка ядра.

Предлагаемые методы реализации

Операцию свертки на базе алгоритма `kn2row` можно разделить на несколько этапов: загрузка данных, матричное перемножение, сдвиг, пропуск и сложение. Последние 3 операции могут выполняться на одном модуле как с использованием буфера (`kn2row-as`), так и без (`kn2row-aa`). Таким образом, можно считать, что алгоритм требует 3 блока в своей лучшей оптимизации. Его преимущество состоит в том, что с увеличением количества каналов у фильтров или самого количества фильтров возрастает его эффективность. Также в нём избегается копирования данных перед перемножением, как это было в `im2`

алгоритмах, из-за этого он считается более эффективным с точки зрения расхода дополнительной памяти. Однако, рассмотренные базовые версии не подразумевают шаг больше 1 и четный размер ядра, поскольку при расчете сдвигов за начало отсчета берется его центр.

Сравнение двух алгоритмов

В таблице 1 приведены затраты дополнительной памяти для разных версий анализируемых алгоритмов. Алгоритм kn2row-aa позволяет существенно снизить показатели, что делает его эффективным с точки зрения памяти. Для im2row, kn2col значения аналогичны.

Таблица 1. Необходимая дополнительная память для рассмотренных алгоритмов [6]

| Аббревиатура | Дополнительная память |
|--------------|-----------------------|
| im2col | (K2-1)(HW)C |
| kn2row | (K2-1)(M(HW)) |
| kn2row-as | ((HW)M) |
| kn2row-aa | KW |

Однако, принимая во внимание критерий универсальности, ситуация будет противоположной. Из-за того, что первый тип алгоритмов преобразовывает изображение для получения окон, то он позволяет осуществлять свертку с более необычными параметрами, например, с нестандартным шагом, большим 1 или произвольным размером ядра. Если оба параметра могут динамично меняться, то имплементация будет подразумевать метод crossbar. В ином случае, можно реализовать этот блок статично [8]. Алгоритмы типа kn2 не позволяют свободно параметризовать шаг, а существенное увеличение размеров ядра приводит к снижению производительности [6].

Что касается временных затрат, то согласно проведенному исследованию на таких процессорах как AMD Ryzen 9 5950X (x86_64), Amlogic S922X Cortex-A53 (ARMv8-A), Intel Celeron 1007U (x86_64), Intel Core i7 930 (x86_32), Intel Core i3 8109U (x86_64), AMD E-450 APU (x86_64) [9] выведены некоторые тенденции при использовании того или иного алгоритма. С увеличением количества каналов входного изображения и ядра (параметр Cout в источнике), временные затраты алгоритма kn2row-aa обычно уменьшаются, поскольку ему не требуется стадия формирования окон, как это происходит перед матричным умножением в im2col. Увеличением размера входного изображения (параметр S в источнике) даёт потенциал алгоритму kn2row-aa при работе с более малым количеством каналов. Это можно объяснить тем, что время на выделение патчей из изображения для алгоритма im2col перед матричным умножением увеличиваются и могут снизить его эффективность настолько, что временные затраты на

получение свертки сравниваются с kn2row-aa или даже превысят его показатели.

Заключение

Каждый из алгоритмов имеет свои преимущества и недостатки. Поэтому необходимо сначала проанализировать предметную область и оценить, какие критерии являются наиболее важными. Также имеет значение техническое оснащение. Для задачи нашей мастерской оба алгоритма могут быть использованы, поскольку каждый требует одинаковое количество этапов и имеет равноценное количество преимуществ и недостатков. На основе проведенного исследования в дальнейшем будет разработан оптимальный вариант использования алгоритмов, в зависимости от поступающих данных.

Список литературы

1. Simard P.Y., Steinkraus D., Platt J.C. Best practices for convolutional neural networks applied to visual document analysis // Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. IEEE Comput. Soc. P. 958–963.
2. Kumar Chellapilla S.P.P.Y.S. High Performance Convolutional Neural Networks for Document Processing // In Tenth International Workshop on Frontiers in Handwriting Recognition. Suvisoft, 2006.
3. Intel Math Kernel Library [Electronic resource]. URL: <http://www.intel.com/cd/software/products/asmona/eng/perflib/mkl/index.html> (accessed: 03.01.2024).
4. AMD Core Math Library (ACML) [Electronic resource]. URL: <https://www.physics.udel.edu/~bnikolic/QTG/shared/docs/acml.pdf> (accessed: 12.02.2024).
5. Automatically Tuned Linear Algebra Software (ATLAS) [Electronic resource]. URL: <http://math-atlas.sourceforge.net/> (accessed: 25.01.2024).
6. Anderson A. et al. Low-memory GEMM-based convolution algorithms for deep neural networks. 2017.
7. Zhou Y. et al. Characterizing and Demystifying the Implicit Convolution Algorithm on Commercial Matrix-Multiplication Accelerators. 2021.
8. Manasi S.D. et al. Reusing GEMM Hardware for Efficient Execution of Depthwise Separable Convolution on ASIC-Based DNN Accelerators // Proceedings of the 28th Asia and South Pacific Design Automation Conference. New York, NY, USA: ACM, 2023. P. 475–482.
9. Trusov A. V. et al. p-im2col: Simple Yet Efficient Convolution Algorithm With Flexibly Controlled Memory Overhead // IEEE Access. 2021. Vol. 9. P. 168162–168184.

АППАРАТНАЯ РЕАЛИЗАЦИЯ СИСТОЛИЧЕСКОГО МАССИВА ДЛЯ РАСЧЕТА МАТРИЧНОГО УМНОЖЕНИЯ

*Ильянова Д.С., Минеев А.В., Рожин А.К.
Национальный исследовательский университет
«Высшая школа экономики»,
департамент электронной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

Систолические массивы являются мощной аппаратной архитектурой для нейронных сетей. Данные структуры обеспечивают значительное повышение производительности и эффективности. Структурированные массивы вычислительных элементов, соединенных между собой в виде конвейера, обеспечивающего эффективный поток данных и параллельную обработку. Они особенно хорошо подходят для реализации сверточных нейронных сетей (CNN) благодаря своей способности выполнять операции свертки с высокой степенью распараллеливания [1]. Это делает их незаменимыми в таких приложениях, как распознавание образов, обнаружение объектов и обработка естественного языка, где CNN играют решающую роль. Аналогия с кровотоком в биологическом сердце помогает наглядно представить их работу: данные ритмично постукают из памяти, проходя через множество элементов, прежде чем вернуться в память [2]. В данной работе представлен результат разработки систолического массива, позволяющий эффективно выполнять матричное умножение на большом количестве узлов.

Введение

В сфере высокопроизводительных вычислений систолические массивы стали мощной и эффективной архитектурой для ускорения широкого спектра вычислительных задач [3]. Возникнув в области параллельной обработки, систолические массивы предлагают структурированный и высокопараллельный подход к обработке данных, что делает их особенно подходящими для приложений, требующих интенсивных вычислений, таких как обработка сигналов, матричное умножение и нейронные сети. Одно из ключевых преимуществ систолических массивов заключается в их способности использовать параллелизм [4]. Это позволяет систолическим массивам достигать высокой вычислительной эффективности и масштабируемости производительности, что делает их пригодными для широкого спектра приложений в различных областях.

В данной работе описывается архитектура разработанного систолического массива в среде разработки Vivado на языке описания аппаратуры System Verilog [5]. Конечный модуль содержит в себе две структуры: модуль МАС-ячейки — главный вычислительный модуль; модуль систолического массива, обеспечивающий корректное соединение матрицы мас-ячеек для обмена информацией между ними.

Приведен список активных ресурсов, которые необходимы для организации систолического массива.

Решение

Модуль систолического массива создан в среде разработки Vivado на языке описания аппаратуры System Verilog. Конечный модуль содержит в себе две структуры: модуль мас-ячейки- главный вычислительный модуль, и модуль систолического массива- модуль, обеспечивающий корректное соединение матрицы мас-ячеек для обмена информации между ними.

Систолический массив

Систолический массив состоит из параметризованного количества идентичных обрабатывающих элементов — МАС-ячеек, которые расположены в виде регулярной сетки (матрицы). Каждая ячейка способна выполнять простую операцию со своими входными данными и передавать результат соседним элементам.

На периферийные МАС-ячейки регулярно подается поток данных, а управление всем систолическим массивом производится по конвейерному принципу. Другими словами, введенный поток данных, проходя все ступени конвейера, обрабатывается каждой МАС-ячейкой и после обработки сбрасывается. По окончании работы систолического результата считывается с регистров МАС-ячеек и передается в виде вектора в выходной регистр. Таким образом, каждая ячейка выполняет операции, ритмично чередуя прием и передачу данных в ходе обмена с окружающими ее ячейками.

Систолический массив представляет из себя параметризуемую структуру: height/width — количество строк/столбцов систолического массива, bits_number — количество бит, выделяемое для записи числа. Для корректной работы систолического массива необходимо указывать параметр k, отвечающий за совпадающее количество строк/столбцов в размерностях перемножаемых матриц. Модуль систолического массива рассчитан на перемножение матриц, содержащих числа типа int8, размером 32×32 .

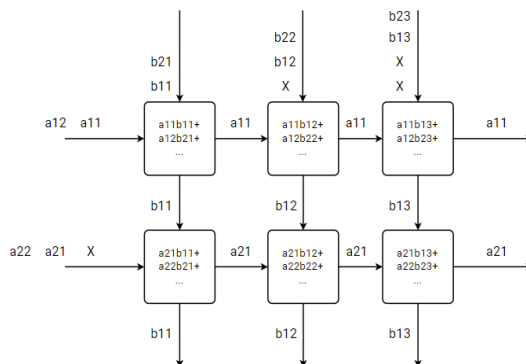


Рис.1. Структура систолического массива

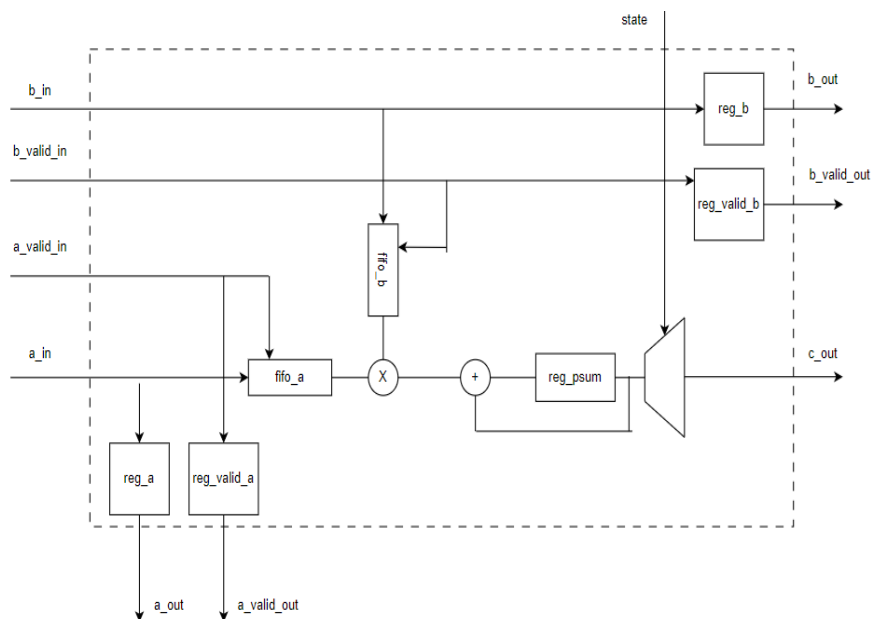


Рис.2. Структура MAC-ячейки

Как показано на рис.2, на вход модуля передается два аргумента для вычисления и сопутствующие им сигналы валидности — управляющие сигналы, определяющие актуальность входных данных. Полученные данные одновременно передаются в буфер и на выход для обеспечения конвейерного продвижения данных по систолическому массиву и на регистры FIFO [6]. Регистры настроены таким образом, что запись на них производится по положительному сигналу валидности, а считывание — в момент, когда оба регистра пусты. Таким образом, регистры FIFO позволяют организовать автономно регулируемые задержки, необходимые для выполнения корректных вычислений. Иначе приходилось бы обеспечивать эти задержки, разбивая поток входных данных незначительными нулями.

После прохождения через регистр FIFO данные попадают в цикл, управляемый конечным автоматом, включающим в себя три состояния: waiting, assum и done. Автомат находится в состоянии waiting до тех пор, пока на него не поступит пара валидных данных, после чего переходит в состояние assum, производит перемножение и сложение поступивших аргументов с последующими данными, а накопленное значение сохраняет в регистр промежуточной суммы, и возвращается обратно в состояние waiting до поступления следующих актуальных данных. Описанный алгоритм повторяется до тех пор, пока количество переходов из состояния waiting в состояние assum не станет равным $k-1$, где k — число, равное совпадающему количеству строк/столбцов в размерностях перемножаемых матриц, затем конечный автомат принимает состояние done.

В этот момент считается, что расчет одного из элементов выходной матрицы окончен. Значение, соответствующее этому элементу, хранится в регистре промежуточной суммы до тех пор, пока на ячейку не будет подан сигнал reset.

MAC-ячейка настраивается по параметру `bits_number` — количество бит, выделяемое для записи числа и устойчива к переполнению.

Моделирование и оценка производительности

Был проведен анализ используемых ресурсов синтезированного систолического массива размером 32×32 . В таблице 1 приведены используемые ресурсы.

Таблица 1. Используемые ресурсы для систолического массива

| Resource | LUT | LUT RAM | FF | BUFG |
|-----------------|--------|-----------|--------|---------|
| Estimation | 58760 | 64 | 67866 | 10 |
| Available | 203800 | 6400 0 | 407600 | 400 |
| Utilization (%) | 28.83 | 0.10 | 16.65 | 4105.25 |

Систолический массив демонстрирует эффективное использование ресурсов: около 28,83% доступных таблиц Look-Up Tables (LUT) отведено для реализации важнейшей комбинационной логики. LUT играют фундаментальную роль в обработке данных внутри массива, способствуя вычислениям с высокой пропускной способностью. Наряду с LUTs, использование LUTRAM, формы памяти на кристалле для хранения данных в LUTs, остается минимальным — 0,10%, что свидетельствует об эффективном использовании и управлении памятью в массиве.

Для хранения промежуточных результатов и управляющих сигналов используются флип-флопы (FF), причем в систолическом массиве задействовано около 16,65 % доступных флип-флопов. Эти FF способствуют эффективной обработке потока данных и операций обработки в структуре массива.

Однако, несмотря на эффективное использование LUT и FF, систолический массив сталкивается с проблемой ресурсов BUFG (Clock Buffers). Массив требует 4105,25% доступных BUFG, что значительно превышает имеющийся запас. BUFG играют важнейшую роль в распределении тактовых сигналов по всей конструкции, обеспечивая синхронную работу и корректность синхронизации.

Профиль использования ресурсов подчеркивает эффективное использование LUT и FF в систолическом массиве, что отражает его способность выполнять сложные вычисления с оптимизированным использованием логики и памяти. Тем не менее, непропорционально высокая потребность в BUFG подчеркивает необходимость дальнейших стратегий оптимизации для балансировки использования ресурсов и повышения общей производительности архитектуры систолического массива.

Заключение

Разработанное устройство является эффективной архитектурой для выполнения операции матричного умножения.

Список литературы

1. Jing Shen, Haoqi Ren, Zhifeng Zhang, Jun Wu, Wenqi Pan, Zhenyu Jiang A High-Performance Systolic Array Accelerator Dedicated for CNN // International Conference on Communication Technology (ICCT). - Xi'an: IEEE, 2019. - С. 1200-1204.
2. Parallel processing – systolic arrays // Geeks for geeks URL: <https://www.geeksforgeeks.org/parallel-processing-systolic-arrays/> (дата обращения: 24.02.2024).
3. Xuechao Wei, Cody Hao Yu, Peng Zhang, Youxiang Chen, Yuxin Wang, Han Hu, Yun Liang, Jason Cong Automated Systolic Array Architecture Synthesis for High Throughput CNN Inference on FPGAs // DAC Design Automation Conference. - 2017. - №54. - С. 1-6.
4. Jai-Hoon Chung, Hyunsoo Yoon, Seung Ryoul Maeng A systolic array exploiting the inherent parallelisms of artificial neural networks // Microprocessing and Microprogramming. - 1992. - №33. - С. 145-159.
5. IEEE Computer Society and the IEEE Standards Association Corporate Advisory Group // IEEE Standard for SystemVerilog— Unified Hardware Design, Specification, and Verification Language URL: https://fpga.mit.edu/6205/_static/F23/documentation/1800-2017.pdf (дата обращения: 24.02.2024).
6. FIFO Generator v13.1. LogiCORE IP Product Guide // Vivado Design Suite URL: <https://docs.xilinx.com/v/u/13.1-English/pg057-fifo-generator> (дата обращения: 24.02.2024).

СОДЕРЖАНИЕ

| | |
|---|-------------|
| <u>Секция «Аналитика, статистика и искусственный интеллект»</u> | 5-60 |
| Федотов Г.А. Анализ алгоритмов обнаружения дипфейков | 5-7 |
| Коровин Е.А. Чиглинцева С.А. Разработка системы распознавания эмоций | 7-11 |
| Тарадайко Е.А. Тузов Н.А. Способность распознавания системы голосового человеко-машинного взаимодействия в условиях воздействия шумов и помех | 11-14 |
| Гурковский А.А. Анализ эффективности алгоритмов согласования в задачах федеративного децентрализованного обучения для сетей с ненадежными узлами | 14-17 |
| Маркелов М.С. Сорочайкин А.Я. Подгорный Л.Е. Лоскутов М.И. Разработка системы предсказания параметров цифровых схем с использованием методов машинного обучения | 17-19 |
| Кожемякин Д.А. Статистический арбитраж на сверхкраткосрочном горизонте с использованием показателя Херста | 19-22 |
| Горбунцов А.А. Миркин Д.Е. Обзор моделей нейронных сетей для улучшения качества речи | 22-25 |
| Горбунцов А.А. Васильев Г.В. Автоматическое определение типа пластика для 3-д печати при помощи нейросетевых технологий | 25-28 |
| Кочарян Д.Р. Использование нейросетевого анализа цифровых рентгенологических исследований органов грудной клетки (флюорографии) для диагностики патологий лёгких с помощью методов компьютерного зрения | 29-31 |
| Чуйко Д.О. Кретьова А.А. Исследование возможностей добавления интерактивных элементов в виртуальные пространства для проведения встреч или корпоративных мероприятий и их влияние на пользователей | 31-33 |
| Блюдова В.М. Разработка методологии оценивания степени суггестивности русскоязычных рекламных текстов | 34-36 |
| Башинский Д.О. Шин В.А. Семичаснов И.В. Исследование и применение технологии Volume Rendering для визуализации результатов МРТ | 37-40 |
| Жариков Е.И. Ключевые аспекты оценки качества систем генерации с дополненной выборкой | 40-43 |
| Захаров Ф.А. Молоканов Р.А. Литвиненко Е.И. Малышкина К.А. Предсказание коммерческой эффективности видеоигр на основе анализа комментариев методами обработки естественного языка | 43-46 |

| | |
|--|---------------|
| Просвирнин Р.О. Дедуктивно-нейронный подход на примере формализации решения задач планиметрии | 46-49 |
| Ковалев Р.А. Ляпунова С.А. Маркин А.А. Попов Л.Д. Модуль визуального программирования для редактора материалов в игровых движках | 49-51 |
| Макшанчиков В.С. Романов И.А. Казаков Л.К. Матанов К.А. Процедурная генерация трехмерной местности в виртуальном пространстве | 51-54 |
| Сиротинский Н.В. Горбунов И.И. Дусов М.А. Яковлев М.В. Якубов В.Ю. Синтез речи с заданными параметрами | 54-57 |
| Прокудина К.В. Скрипленок М.С. Разработка детектора печати на изображениях | 57-60 |
| <u>Секция «Компьютерное моделирование физико-механических систем»</u> | 61-80 |
| Анохин В.В. Миколаенко В.В. Кудряшов М.Д. Автоматизация построения модели деформационного поведения сверхпластичных материалов | 61-64 |
| Вараксин И.М. Писарев В.В. Структура данных для выделения кластеров частиц в молекулярном моделировании | 64-65 |
| Толстиков С.Ю. Муллахметов И.Р. Моделирование радиационного заряжения полимерной изоляции проводов космического применения при низких температурах | 66-68 |
| Ершов И.С. Исследование эффективности поиска коллизий в геометрических моделях при разных способах представления объектов | 68-71 |
| Мкртчян Я.Р. Смирнов Г.С. Предсказание вязкости углеводов с помощью молекулярной динамики и машинного обучения | 71-72 |
| Варфоломеева А.А. Паршина С.Р. Пятко Л.А. Моделирование и анализ конструкции микрофлюидного чипа | 73-75 |
| Пиле Я.Э. Влияние связанных состояний на сверхпроводимость в модели Хаббарда на лестничных решетках | 75-78 |
| Недомолкин И.Э. Конилов М.П. Анализ и оценка производительности суперкомпьютера “сHARISMa” и его компонентов для первопринципного моделирования материалов | 78-80 |
| <u>Секция «Математическое моделирование и динамические системы»</u> | 81-107 |
| Рахмаева Е.М. Иванов Г.Д. Оптимизация сетки кредитных предложений | 81-84 |
| Борисов В.Д. Математическое моделирование полевой эмиссии из катода малых размеров | 84-86 |
| Нечитайло В.Г. Аналитические свойства обобщенного уравнения Льюнара | 86-89 |

| | |
|--|----------------|
| Иваненко О.В. Шаяхметова Д.Т. Математическая модель динамики мнений с ограниченным доверием с внешним влиянием | 89-92 |
| Глазунова А.М. Вывод уравнений двухпалубного обтекания неровной поверхности с малыми быстро осциллирующими неровностями | 92-95 |
| Петухова Е.С. Об обосновании неосциллирующих ВКБ асимптотик | 95-97 |
| Шумов Н.А. Применение полумарковских процессов для моделирования системы массового обслуживания $G G n$ | 97-98 |
| Гафурова Д.Р. Исследование периодических решений в окрестности точек либрации в круговой ограниченной задаче трех тел | 98-103 |
| Посынкин А.А. Протасов Т.Е. Применение методов машинного обучения для вычисления орбит вокруг точек либрации в круговой ограниченной задаче трех тел | 103-106 |
| Гаража И.А. Оптимальное управление нелинейным объектом, линеаризуемым обратной связью, в задаче дифференциальной игры | 106-107 |
| <u>Секция «Компьютерные сети, телекоммуникационные системы и интернет вещей»</u> | 108-180 |
| Быковский Н.А. Программное обеспечение для программно-управляемого имитатора тока | 108-110 |
| Рзаев Э.Р. Вычисление ширины бисекции циркулянтного графа | 110-112 |
| Лясковский А.Д. Система позиционирования внутри помещений для платформы интернета вещей | 112-115 |
| Бахшалиев Р.М. Мерзлинкин В.Е. Наземная оптическая приемная система для квантового канала связи спутник-Земля | 115-118 |
| Желтова Д.Н. Тестирование аппаратной части транспортного робота с интеллектуальной системой управления | 118-119 |
| Чашкин Л.Б. Подкопаева П.О. Маршутина Е.Н. Автоматизация обновления карт дорожной сети с использованием GPS-трекера | 119-122 |
| Асташов С.Д. Елисеенко А.М. Разработка SIP системы служебной связи | 122-123 |
| Сафронов А.М. Михайлов К.В. Обзор методов развертывания сетей именованных данных | 124-125 |
| Горовой В.В. Прокудина К.В. Апыюк В.Р. Огай В.А. Реконфигурируемые стенды | 125-128 |

| | |
|---|---------|
| Айрапетян Т.Ж. Пардаев А.А. Шубекова А.Б. Разработка системы виртуальной студии для видеопроизводства в реальном времени | 128-130 |
| Сенников М.В. Елисеенко А.М. Разработка сервиса проксирования ONVIF PTZ | 130-133 |
| Егоров А.М. Подгук А.Е. Половцев А.В. Чертов В.Д. Корректирующие подходы к визуализации геолоцированных объектов маршрутизации на картографических подложках | 133-136 |
| Омельченко К.Г. Стрельников Д.Ю. Применение теории принятия решений в задаче выбора стратегии реагирования на ИТ-инциденты | 136-138 |
| Никитин С.Е. Ключев Н.С. Сахаутдинов Ш.Р. Предложение структуры модели системы связи для высокоскоростных железнодорожных магистралей (ВСЖМ) | 138-141 |
| Севрюков Д.О. Платонов Д.Д. Лазарев М.А. Потапов Д.С. Определение рабочего режима SNSPD для детектирования спутникового телекоммуникационного оптического сигнала | 141-142 |
| Портной С.Л. Никитин С.Е. Волошин А.Д. Антошкин Г.Д. Оценка производительности полярных и LDPC кодов, используемых в сценариях ВСЖМ | 142-145 |
| Эбрахим А. Предметно-ориентированный язык для оптимизации систем Интернета вещей | 145-147 |
| Маркова Э.С. Макагонова В.Д. Нигматуллин Н.Р. Применение нейронной сети для прогнозирования точки насыщения в сетях на кристалле | 147-150 |
| Еременко Е.А. Лыжин И.Г. Концепция программно-аппаратной платформы для реализации образовательных программ в сфере Интернета вещей: возможности, применимость и перспективы | 150-153 |
| Фролов К.Д. Мирошниченко Л.И. Исследование реализации криптографического расширения архитектуры RISC-V | 153-155 |
| Тимошкевич И.В. Цифровые двойники на базе развития технологий ВІМ и ІОТ | 155-159 |
| Буслаев Д.П. Ельцов М.Е. Портной С.Л. Исследование арендных решений на российском рынке телекоммуникаций | 159-162 |
| Андрончев А.Д. Насыров Д.Д. Рахимуллина В.Е. Мкртчян Г.А. Бушуев Н.И. Интегрирование 3D модели распространения сигналов в среду моделирования для беспилотного и подключенного транспорта | 162-166 |
| Шембель Д.А. Богомолов Р.Д. Сивов Г.В. Минибаева А.И. Чо Хансли Че Программные модули для расширения симулятора ns-3 и фреймворка MoReV2X | 166-169 |
| Курилович П.А. Обеспечение наблюдаемости систем на основе микросервисной архитектуры | 169-172 |
| Арутюнян А.Р. Осинцев Н.И. Омаров М.Т. Нестеренко А.А. Сравнительный анализ оптимизаций компиляторов Clang и GCC на примере компиляции высоконагруженной системы | 172-176 |

| | |
|--|---------|
| Ковалькова М.Е. Гавриков А.А. Бадалян Р.К. Дуплин Г.В. Эффективная синхронизация работы нескольких пользователей в слайсере | 177-179 |
| Васильев И.А. Фирстова Н.М. Разработка визуализатора данных векторного файла G-Code | 179-180 |
| <u>Секция «Материалы и технологии»</u> | |
| Иштеев Р.Р. Кондратенко В.С. Технология лазерного скрайбирования солнечного элемента на основе перовскита | 181-183 |
| Корнев С.А. Андреев В.В. Определение дефектности подзатворного диэлектрика МДП-структур | 183-184 |
| Венедиктов И.О. Исследование направленных ответвителей на базе интегральной оптики на нитрид кремниевой платформе | 185-187 |
| Епифанов Н.А. Дёмин А.С. Морозов Е.В. Влияние пучково-плазменного облучения на поверхность ванадия и выброс частиц в установке Плазменный фокус | 187-190 |
| Ручий А.Ю. Домарев С.Н. Рыжих Д.А. Пушкарская Д.В. Модификация поверхности полимеров посредством воздействия на них вакуумного ультрафиолета | 190-192 |
| Дёмин А.С. Епифанов Н.А. Морозов Е.В. Сравнительное исследование повреждаемости поверхностного слоя пластин тантала при импульсном воздействии потоков быстрых ионов и высокотемпературной плотной плазмы гелия и дейтерия в установках «Плазменный фокус» с различной энергетикой | 192-195 |
| Морозов Е.В. Дёмин А.С. Епифанов Н.А. Воздействие импульсных потоков высокотемпературной плазмы и быстрых ионов на никелевый сплав Инконель 718 | 195-198 |
| Галияхмедов Э.И. Гончаров Н.П. Герасимов Р.А. Нгуен А. Тупицкий Д.С. Разработка параметризованного 3D принтера | 198-200 |
| Ерофеева А.Р. Иким М.И. Структура, проводимость и сенсорные свойства гидротермальных композитов ZnO-In ₂ O ₃ | 200-202 |
| <u>Секция «Низкоразмерные квантовые системы»</u> | |
| Гурский А.С. Экспериментальное исследование неравновесного квазичастичного транспорта в сверхпроводнике | 203-205 |
| Галиев А.Ф. Карамов Д.Д. Каримов В.Р. Ишмухаметов М.С. Буланкин Н.С. Размерные эффекты в субмикронных пленках полиарилефталидов | 206-207 |
| Сеидов С.С. Селезнёв Д.В. Пугач Н.Г. Решение уравнений Узаделя для контакта ферромагнитный изолятор-сверхпроводник-ферромагнитный изолятор в линейном приближении | 208-209 |
| Смазнова Х.Т. Арутюнов К.Ю. Завьялов В.В. Юсупов А.Р. Карамов Д.Д. Лачинов А.Н. Электрическая проводимость слоистых структур индий-полидефинеленфталид-индий при низких температурах | 210-212 |

| | |
|--|----------------|
| Туркин Я.В. Пугач Н.Г. Динамический эффект близости в гибридной структуре сверхпроводник ферромагнитный диэлектрик | 212-214 |
| Безмянных Д.Г. Пугач Н.Г. Седов Е.Г. Арутюнов К.Ю. Исследование влияния квантово-размерного эффекта на сверхпроводящий переход в тонких пленках алюминия | 214-215 |
| Лекомцев Н.В. Пугач Н.Г. Управляемый эффект близости в структуре сверхпроводник-геликоидальный магнетик | 216-218 |
| Селезнёв Д.В. Пугач Н.Г. Особенности поведения плотности состояний на границах гетероструктуры сверхпроводник ферромагнитный изолятор | 218-220 |
| Милютин Д.П. Пугач Н.Г. Спиновое рассеяние в гетероструктурах сверхпроводник-ферромагнетик | 221-222 |
| Позднякова Е.Ф. Арутюнов К.Ю. Тарасов М.А. Чекушкин А.М. Маркина М.А. Методика изготовления и физические свойства туннельных переходов микронных размеров | 223-225 |
| Пашковская В.Д. Моделирование сверхпроводящих тонких пленок с использованием нестационарных уравнений Гинзбурга-Ландау с помощью переменных связки | 225-227 |
| <u>Секция «Функциональные квантовые материалы»</u> | 228-245 |
| Амиров Э.Ш. Васенко А.С. Исследование неравновесных процессов в туннельных контактах металл-сверхпроводник | 228-231 |
| Кононова Н.А. Моделирование нанопроволок с использованием нестационарных уравнений Гинзбурга-Ландау | 231-233 |
| Тюменев Р. Калашников Д.С. Шишкин А.Г. Столяров В.С. СВЧ-генератор на основе джозефсоновского перехода | 234-235 |
| Тейшейра Сарайва Т. Диссипация и множественные характерные длины во время динамических событий в грязных сверхпроводящих тонких пленках | 235-236 |
| Саматов М.Р. Изучение атомных свойств на межзеренной границе в перовскитных материалах | 237-239 |
| Глуховцев П.И. Никонов Э.Г. Компьютерное исследование природы фазовых переходов в системах заряженных частиц, ограниченных потенциалом с круговой симметрией | 239-242 |
| Абзах Б.А. Васенко А.С. Исследование влияния структурного беспорядка на ширину запрещенной зоны галоидного перовскита Cs ₂ AgBiBr ₆ | 242-244 |
| Карабасов Т. Васенко А.С. Сверхпроводящий диодный эффект в топологических гибридных структурах | 244-245 |

| | |
|--|----------------|
| <u>Секция «Кибербезопасность»</u> | 246-292 |
| Сербаев В.Р. Принцип работы деревьев Веркла и исследование возможностей их применения для решения задач защиты информации | 246-249 |
| Сахаров К.Е. Методы обхода систем обнаружения/предотвращения вторжений и способы защиты от них | 249-251 |
| Краюшкин Д.В. Центр управления и мониторинга компонентами радиологической информационной системы | 251-253 |
| Гришкина В.И. Исследование угроз и рисков информационной безопасности финансово-аналитической системы НИУ ВШЭ | 254-257 |
| Жидков А.С. Распознавание аудио контента по данным акселерометра | 257-260 |
| Мочалов И.С. Уязвимости информации при передаче автоинкрементных идентификаторов полей баз данных через REST API | 260-264 |
| Вербенко С.С. Сравнение протоколов передачи данных WebSocket и HTTP/2.0 | 264-266 |
| Чибрин М.Д. Исследование возможности применения нейронных сетей к построению универсальных атак на хеш функции | 266-269 |
| Шабаев С.П. Исследование генерации случайных последовательностей с использованием цепей Маркова | 269-272 |
| Попов В.А. Сравнение моделей информационного воздействия в сети Telegram-каналов | 272-274 |
| Вишнёв Е.А. Исследование по автоматической генерации позитивной модели безопасности для ModSecurity WAF | 275-277 |
| Сарибекян Г.Э. Валиев Ш.Т. Манжилевский С.С. Простенко А.А. Стандартизация постквантовых цифровых подписей | 278-281 |
| Сарибекян Г.Э. Якушенокс К.П. Пашина К.И. Коробков В.С. Схемы обмена ключами в постквантовой криптографии на основе QC-MDPC-кодов | 282-285 |
| Павлова Е.Д. Валиев Ш.Т. Коробков В.С. Манжилевский С.С. Исследование подходов к организации скрытых вычислений | 285-288 |
| Налегач Д.И. Старых В.А. Обеспечение информационной безопасности при проектировании и документировании API с использованием OpenAPI и Spectral | 289-292 |

**Секция «Микроволновая и оптическая электроника,
электромагнитная совместимость»** **293-322**

- Львов А.В. Ивашенцева И.В. Бондарева П.И. Панарина П.В.** 293-295
Программируемые отражающие поверхности для беспроводной связи нового поколения
- Машкова М.А.** 295-298
Исследование волноводного поляризатора с клиновидной магнитной стенкой
- Посмитный С.В. Ботов М.А.** 298-299
Установка для измерения показателя преломления
- Сетдиков Р.М. Тюрин И.Г.** 299-301
Разработка автономного гамма-спектрометра: решение основных задач
- Воронова Е.И. Скрябина В.А.** 301-305
Моделирование фрактальных антенн Коха и Пеано с двухслойной мета-поверхностью
- Левашов С.Д.** 305-307
Разработка гиперспектрального оптического модуля для БПЛА в задачах атмосферных исследований, валидации данных ДЗЗ и оптической связи в открытом пространстве
- Цуркан А.А. Антонов М.А. Коробок М.А. Дуплин Г.В.** 308-310
Система сбора биопотенциалов мышц у людей с ампутацией верхней конечности для воспроизведения мелкой моторики
- Ильянова Д.С. Кацнельсон А.И. Колпаков М.М.** 311-313
Позиционирование человека в пространстве микроволновым излучением на основе метода Доплера
- Андреева А.А. Уткин Б.В.** 313-316
Исследование спектральных характеристик контактных радиопомех для решения задачи обеспечения электромагнитной совместимости на мобильных объектах
- Денисов М.В.** 316-320
Математическая модель расчета и методика измерений параметров анизотропного магниторезистивного датчика углового перемещения
- Носов И.А. Игнатьев Д.К. Янцен Е.Э.** 320-322
Разработка усилителя биоэлектрических потенциалов с быстрым мультиплексированием электродов для портативной системы регистрации поверхностной электромиограммы высокой плотности

Секция «Электронное приборостроение» **323-408**

- Козаев З.З. Посмитный С.В. Старилова Е.В.** 323-325
Система автоматической замены аккумулятора БПЛА
- Елькин В.А. Мотайленко И.А. Чернявских И.И. Носов И.А.** 325-327
Приём, передача и расшифровка сообщений с передатчиков ADS-B с помощью малых космических аппаратов
- Костюк А.А. Старкова Э.А.** 327-331
Математическая модель расчётной оценки показателей безотказности информационной системы наземного комплекса управления роём беспилотных летательных аппаратов

| | |
|---|---------|
| Зубкова А.И. Харитонов И.А. Реализация электро-тепловых расчетов силовых схем в Python Spice с использованием цифровых двойников мощных МОП транзисторов | 332-334 |
| Ягжов А.А. Вершков М.Д. Романов Н.С. Знатнов Е.П. Федотова А.А. Экспериментальное сравнение методов синтеза логических схем | 335-338 |
| Купцова Д.Д. Кучина А.П. Старкова Э.А. Программно-аппаратная реализация методики ускоренных испытаний маломощных резисторов | 338-341 |
| Чуча Г.С. Исмаил-Заде М.Р. Исследование и моделирование статистического разброса параметров JFET-транзисторов с учетом влияния температуры | 341-343 |
| Бадалян Р.К. Ключев Н.С. Мотайленко И.А. Цуркан А.А. Разработка портативного экструдера для производства филамента | 343-346 |
| Алешина А.И. Буцкий Д.О. Чиглинцева С.А. Структура системы оценки показателей зрения человека | 347-349 |
| Грицун Т.Ю. Кашинцев И.К. Портнягин Е.Р. Ржевская М.М. Чижмар М.М. Разработка инструментов верификации САПР цифровых СБИС | 350-352 |
| Стрельцов Г.А. Юлдашов А.С. Алгоритм обратной связи рулевого управления для VR-симулятора яхты | 352-356 |
| Бушуев Н.И. Козлова А.А. Мкртчян Г.А. Насыров Д.Д. Модернизация структуры опросника для оценки коэффициента качества производства электронных модулей на основе справочника “Handbook of 217plus reliability prediction models” | 356-359 |
| Цветков В.Э. Ландер Л.Б. Анализ методов оценки надежности систем промышленного интернета вещей | 359-361 |
| Мешочков Д.М. Ушаков А.Н. Инструменты трансляции HDL-описаний для использования в САПР | 361-364 |
| Цуркан А.А. Волошин А.Д. Ключев Н.С. Коробок М.А. Сушильная камера, как способ уменьшения брака при печати на 3D принтере | 364-367 |
| Костиков Д.В. Кацнельсон А.И. Методика определения базовых показателей надежности ЭКБ методами машинного обучения | 367-369 |
| Щукин А.А. Рожин А.К. Обзор технологий edge AI, применяемых в автоиндустрии | 370-371 |
| Губазов Д.Д. Кателкин Н.И. Коммуникация процессоров в гетерогенной вычислительной системе | 371-374 |
| Рахимов Р.М. Анализ аппаратных методов автоматической оптической инспекции печатных плат и узлов | 375-376 |
| Шарапов А.Р. Нечеухина А.А. Исследование механических и тепловых режимов электронных приборов управления электромеханическим приводом рулевого двигателя космического корабля | 377-379 |

| | |
|--|---------|
| Козаев З.З. Чуча Г.С. Разработка макромоделей аналоговых интегральных микросхем для схемотехнического моделирования аппаратуры | 379-381 |
| Волков П.О. Шайкин К.А. Ясафов А.В. Алюшина Е.Р. Обзор программных продуктов по оценке надежности технических средств | 381-384 |
| Старилова Е.В. Хобов А.П. Исмаил-Заде М.Р. SPICE моделирование элементов КНС БИС с учетом технологического разброса | 384-386 |
| Пугач В.П. Чернявских И.И. Индюченко Н.А. Гульев А.В. Открытая система хранения и обработки набора данных комбинационных схем | 387-389 |
| Клопотов Г.И. Ивлев К.И. Шундеева А.А. Автоматизация процесса моделирования работы КМОП схем с учетом воздействия на них внутренних факторов старения со временем | 389-392 |
| Кацнельсон А.И. Кофанова М.А. Новиков К.В. Обзор генеративных языковых моделей для проектирования печатного узла | 392-395 |
| Дубленский Д.С. Махмудов Т.Н. Обзор датчиков полётного контроллера беспилотного летательного аппарата | 395-397 |
| Сазыкин Д.Н. Данилов Е.М. Лебедев М.О. Ключев Н.С. Анализ кинематических систем для FDM 3D печати: выбор эффективного решения для производства прототипов и функциональных деталей | 397-401 |
| Кузнецов А.Г. Тумковский С.Р. Пожидаев Е.Д. Исследование воздействия электростатических разрядов на мощные МОП транзисторы с учётом влияния ёмкости печатной платы | 401-403 |
| Деркач В.В. Батонова О.Ю. Рожин А.К. Обзор основанных на матричном умножении алгоритмов свёртки на аппаратном уровне | 403-405 |
| Ильянова Д.С. Минеев А.В. Рожин А.К. Аппаратная реализация систолического массива для расчета матричного умножения | 406-408 |
| СОДЕРЖАНИЕ | 409-418 |

**Межвузовская научно-техническая конференция студентов,
аспирантов и молодых специалистов имени Е.В. Арменского.
Материалы конференции.**



Подписано в печать 30.10.2024 г. Формат 60x84/8. Бумага офсетная №2.
Печать ризография. Усл.печ.л. 52,25. Уч.-изд.л. 47,03. Тираж 100 экз.
ООО "Издательский дом Медиа Паблшер".