Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых» (ВлГУ)

На правах рукописи

Борданов Илья Алексеевич

МОДЕЛИ И АЛГОРИТМЫ ОЦЕНКИ ФУНКЦИОНАЛЬНОЙ КОРРЕКТНОСТИ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ НА БАЗЕ МЕМРИСТОРОВ

Специальность 2.3.1 – «Системный анализ, управление и обработка информации, статистика»

Диссертационная работа на соискание ученой степени кандидата технических наук

Научный руководитель
Сергей Андреевич Щаников, к.т.н., доцент

Владимир, 2026 год

ОГЛАВЛЕНИЕ

Введение 4

1 Современные методы оценки и обеспечения функциональной корректности искусственных нейронных сетей на базе мемристоров 15

1.1 Особенности аппаратной реализации искусственных нейронных сетей 15

1.2 Факторы, влияющие на функциональную корректность искусственных нейронных сетей на базе мемристоров 19

1.3 Методы обеспечения функциональной корректности искусственных нейронных сетей на базе мемристоров 21

1.4 Методы оценки функциональной корректности искусственных нейронных сетей на базе мемристоров 27

1.5 Обоснование разработки новых моделей и алгоритмов для оценки функциональной корректности искусственных нейронных сетей на базе мемристоров 33

1.6 Выводы по главе 36

2 Разработка моделей и алгоритмов, позволяющих установить взаимосвязь между параметрами сигналов задания сопротивления и функциональной корректностью искусственных нейронных сетей на базе мемристоров 39

2.1 Модель и алгоритм моделирования зависимости сопротивления мемристивного устройства от параметров сигналов его задания 39

2.2 Модель и алгоритм моделирования зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса 42

2.3 Алгоритм оценки функциональной корректности искусственных нейронных сетей на базе мемристоров 45

2.4 Исследование предложенных моделей и алгоритмов моделирования 48

2.7 Выводы по главе 63

3 Программно-аппаратный комплекс для оценки функциональной корректности искусственных нейронных сетей на базе мемристоров 65

3.1 Описание исследуемых металл-оксидных мемристивных устройств 65

3.2 Описание аппаратной части программно-аппаратного комплекса 66

3.3 Описание программной части программно-аппаратного комплекса 72

3.4 Выводы по главе 80

4 Практическое применение разработанных моделей и алгоритмов, аппаратных и программных средств 82

4.1 Построение модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания 82

4.2 Построение модели зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса 86

4.3 Оценка функциональной корректности тестовых искусственных нейронных сетей на базе металл-оксидных мемристивных устройств 89

4.4 Выводы по главе 98

Заключение 101

Список сокращений и условных обозначений 103

Список литературы 105

Приложение А 121

Приложение Б 124

ВВЕДЕНИЕ

Актуальность темы исследования

Мемристивное устройство (МУ) или мемристор – это пассивный элемент в микроэлектронике, сопротивление которого изменяется под действием электрического поля и протекавшего через него заряда, и сохраняется длительное время. В искусственных нейронных сетях (ИНС) МУ используются для выполнения операции матричного умножения, которое происходит в соответствии с законами Ома и Кирхгофа. Поскольку вычисления аналоговые, а МУ имеют вариации сопротивлений от устройства к устройству и от цикла записи к циклу, то для того, чтобы создавать нейропроцессоры на базе МУ, важно уметь оценивать функциональную корректность (ФК) ИНС на базе мемристоров (ИНСМ) на этапе проектирования [1,2]. Метрики оценки ФК ИНС подлежат обязательному расчету при оценке качества систем искусственного интеллекта (ИИ) в соответствии с ГОСТ Р 59898-2021.

В ходе анализа публикаций было установлено, что метрики оценки ФК ИНСМ аналитически рассчитать можно лишь для простых демонстрационных случаев. В настоящее время, для более сложных архитектур в основном применяется компьютерное моделирование, при котором используются различные математические модели МУ и принципы внесения вариаций в их работу.

Одними из наиболее распространенных моделей МУ являются модели, описывающие связь между приложенным напряжением, током в цепи с МУ и переменной состояния (модели вольтамперных характеристик (ВАХ)) [3]. С помощью этих моделей можно описать процесс резистивного переключения в МУ, однако применительно к оценке ФК ИНСМ такие модели имеют ряд недостатков. Эти недостатки заключаются в сложности и неоднозначности при учете и задании вариаций сопротивлений от цикла к циклу и от устройства к устройству, и в высокой ресурсоемкости процесса моделирования ИНСМ, требующего решения

одного или нескольких дифференциальных уравнений для каждого МУ, что критично для больших ИНСМ, состоящих из тысяч или миллионов МУ.

Менее ресурсоемким методом оценки ФК больших ИНСМ является моделирование через задание разбросов весов синапсов нейронов. Однако, при использовании такого метода отсутствует связь с параметрами сигнала задания сопротивления, что усложняет синтез конкретных значений параметров сигналов после анализа ФК ИНСМ. Таким образом, разработка моделей и алгоритмов для оценки ФК ИНСМ на основе взаимосвязи между параметрами сигнала задания сопротивления и получаемыми для конкретных МУ сопротивлениями является актуальной задачей.

Объект исследования – искусственные нейронные сети на базе мемристивных устройств.

Предмет исследования – модели и алгоритмы оценки функциональной корректности искусственных нейронных сетей на базе мемристивных устройств.

Степень разработанности темы исследования

В системный анализ, оценку качества и надежности, а также разработку сложных вычислительных и управляющих систем существенный вклад внесли Каляев И.А., Хранилов В.П., Монахов М.Ю., Полушин П.А., Давыдов Н.Н., Самойлов А.Г., Жизняков А.Л., Ромашов В.В., Веселов О.В. и др.

В разработку различных методов моделирования мемристивных устройств на физическом уровне с учетом погрешностей их функционирования существенный вклад сделали Бутусов Д.Н., Островский В.Ю., Агудов Н.В., Гусейнов Д.В., Roldan J., Serb A., Stathopoulos S., Prodromakis T. и др.

Исследованиями в области оценки влияния этих погрешностей на функциональную корректность ИНСМ на информационном уровне, сделавшими существенный вклад в данную область, занимались Данилин С.Н., Емельянов А.В., Демин В.А., Mehonic A., Yang J., Ielmini D. и др.

Таким образом, в настоящее время существует противоречие, при котором формализация связи между физическим и информационным уровнем моделирования ИНСМ затруднена – с одной стороны качественные модели

отдельных МУ адекватно описывают их работу, но сопряжены со значительными вычислительными затратами при создании моделей больших ИНСМ и нейроморфных вычислителей, с другой стороны модели ИНСМ, учитывающие только разброс весов, не имеют взаимосвязи с процессом записи этих весов. Необходимо применение системного подхода, при котором ИНСМ будет рассматриваться как единая система и будет определяться связь между физическими и информационными процессами в ней.

Цели и задачи

Целью диссертационной работы является формирование новых моделей и алгоритмов для оценки функциональной корректности искусственных нейронных сетей на базе мемристоров, обеспечивающих повышение степени точности результатов моделирования на этапе проектирования.

Для достижения поставленной цели были поставлены следующие **задачи исследования**:

1) Разработка модели и алгоритма моделирования зависимости сопротивления мемристивного устройства от параметров сигналов его задания.

2) Разработка модели и алгоритма моделирования зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса.

3) Разработка алгоритма оценки функциональной корректности ИНСМ для задачи классификации и архитектур сверточных, рекуррентных и полносвязных сетей прямого распространения на основе метрики оценки доли правильных исходов с учетом выбранных параметров сигналов задания сопротивлений мемристивных устройств и параметров реально заданных сопротивлений, схемы формирования веса и максимально допустимых напряжений на выходе нейронов.

Проблематика, исследованная в диссертации, соответствует пунктам 4, 5, 11 паспорта специальности 2.3.1 «Системный анализ, управление и обработка информации, статистика».

Научная новизна

1) Разработана новая модель и алгоритм моделирования зависимости сопротивления мемристивного устройства от параметров сигналов его задания,

отличающиеся тем, что данная зависимость описывает не функциональную взаимосвязь между параметрами сигналов и физическими процессами в мемристоре при прохождении данного сигнала, а статистическую взаимосвязь между параметрами сигнала задания сопротивления и конечным значением сопротивления, и позволяющие рассчитать погрешность задания сопротивления мемристора.

2) Разработана новая модель и алгоритм моделирования зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса, отличающиеся от существующих тем, что вес представляется не аналитической зависимостью между электрическими параметрами цепи, а статистической зависимостью между сопротивлением и весом, и позволяющие рассчитать погрешность задания веса ИНСМ.

3) Разработан оригинальный алгоритм оценки функциональной корректности ИНСМ, отличающийся от существующих тем, что погрешности весов различны для каждого номинального значения веса и определяются из моделей зависимости погрешности веса от сопротивления и параметров сигнала задания сопротивления, а также тем, что в нем учитываются ограничения максимально допустимых рабочих напряжений на входе сети, и позволяющий оценить функциональную корректность ИНСМ максимально приближенно к реальному устройству.

Теоретическая и практическая значимость работы

Теоретическая значимость результатов исследований заключается в том, что разработанные модели и алгоритмы позволяют получать новые знания о процессах задания сопротивлений в МУ и о влиянии погрешностей задания сопротивлений на ФК ИНСМ. Полученные знания о влиянии вариаций сопротивлений МУ на долю правильных исходов ИНСМ для задачи классификации при различных архитектурах, таких как сверхточные, рекуррентные или полносвязные сети прямого распространения, позволяют углубить знания об устойчивости нейроморфных архитектур к данному типу погрешностей. Разработанные алгоритмы и модели оценки ФК обеспечивают

теоретическую базу для повышения степени точности результатов моделирования ИНСМ.

Полученные модели и алгоритмы в перспективе создают основу для последующей разработки требований к качеству изготовления МУ, методов коррекции ошибок ИНСМ и стратегий тестирования при создании нейропроцессоров.

Практическая значимость результатов исследований состоит в том, что разработанные модели и алгоритмы позволяют повысить степень точности результатов оценки ФК ИНСМ с учетом вариаций сопротивлений конкретных мемристивных устройств, а также подобрать значения параметров импульсных сигналов для задания нужных сопротивлений. Результаты вычислительных и натурных испытаний подтверждают это – разница между оценкой доли правильных исходов ИНСМ в модели и эксперименте предложенным методом не превышает 3 %. Разработанное методологическое и программно-аппаратное обеспечение внедрено в научно-исследовательский процесс лаборатории разработки систем искусственного интеллекта (ИИ) МИ ВлГУ, лаборатории мемристорной наноэлектроники НОЦ ФТНС ННГУ им. Н.И. Лобачевского и в производственном процессе ООО «Поликетон» (приложение А).

В перспективе результаты исследования могут быть использованы при проектировании систем интернета вещей и носимой электроники, где одними из ключевых параметров является низкое энергопотребление и возможность обработки данных непосредственно на самом устройстве для защиты информации, что может быть реализовано с помощью нейроморфных систем на безе МУ. Важным архитектурным преимуществом таких систем ИИ является отказ от энергозависимой памяти и возможности выполнения матричного умножения за один такт, что в значительной степени снижает потребление энергии.

Методология и методы исследования

Для решения поставленных задач применены методы системного анализа, теории планирования эксперимента и статистической обработки экспериментальных данных для создания моделей вариации сопротивления МУ и

веса синапса нейрона, методология компьютерного моделирования для оценки влияния вариаций сопротивлений мемристивных устройств на ФК ИНСМ, методы машинного обучения для обучения ИНСМ.

Положения, выносимые на защиту

1) Результаты вычислительных экспериментов применения модели и алгоритма моделирования зависимости сопротивления мемристивного устройства от параметров сигналов его задания подтверждают, что модельные и экспериментальные данные совпадают с доверительной вероятностью 95%.

2) Результаты вычислительных экспериментов применения модели и алгоритма моделирования зависимости веса синапса нейрона и погрешности веса от сопротивления мемристивного устройства и схемы формирования веса подтверждают, что модельные и экспериментальные данные совпадают с доверительной вероятностью 95%.

3) Результаты вычислительных и натурных экспериментов применения алгоритма оценки функциональной корректности ИНСМ подтверждают, что использование предложенного алгоритма позволяет повысить степень точности результатов моделирования. Разница между оценкой долей правильных исходов ИНСМ в модели и эксперименте предложенным алгоритмом не превышает: 1% — для сверточных ИНСМ, 2% — для рекуррентных ИНСМ и 3% — для полносвязных ИНСМ прямого распространения.

Степень достоверности и апробация результатов

Полученные в диссертации результаты моделирования зависимости сопротивлений МУ и весов синапсов нейронов ИНСМ от параметров сигналов их задания согласуются с экспериментальными данными, полученными с помощью реального устройства (кроссбар-массив МУ 32x8 1T1R) (модельные и экспериментальные данные совпадают с доверительной вероятностью 0,95). Результаты компьютерного моделирования ИНСМ и оценки ФК проведены для нескольких разных архитектур ИНС (полносвязная ИНС прямого распространения, сверточная ИНС, рекуррентная ИНС) и практических задач. Проведено сравнение результатов компьютерного моделирования с аппаратно реализованными ИНСМ.

Разница между оценкой доли правильных исходов ИНСМ в эксперименте и в модели не превышает 3 %, в то время как для оценки путем задания разброса весов доходит до 25 %). Методологическое и программно-аппаратное обеспечение разработаны в ходе выполнения следующих НИР: «Разработка и исследование методов имитационного моделирования искусственных нейронных сетей на базе мемристоров на основе теории планирования эксперимента» (грант РФФИ №18-38-00592 мол_а, 2018-2020); «Высокопроизводительные аппаратные ускорители искусственного интеллекта на базе мемристивных устройств» (субсидия Министерства науки и высшего образования РФ, 2021-2022 (проект № 13.2251.21.0098, соглашение № 075-15- 2021-1017)). Полученные результаты использовались для работы с мемристивными устройствами в следующих НИР: «Разработка научно-технологических принципов создания и функционирования нейроморфных систем аналогового машинного зрения на основе мемристивных устройств» (грант РНФ № 21-71-00136, 2021-2023); «Исследование и моделирование механизмов и аналоговых систем векторно-матричного умножения на базе устройств с эффектом резистивного переключения для создания энергоэффективных нейропроцессоров» (стипендия Президента РФ СП-3988.2022.5, 2022-2024); «Нейроморфные системы обработки информации и управления на основе мемристорной наноэлектроники» (Дополнительное соглашение № 075-03-2025-425/1 от 25.03.2025 к Соглашению о предоставлении субсидии из федерального бюджета на финансовое обеспечение выполнения государственного задания на оказание государственных услуг (выполнение работ) от 26.03.2025 г., шифр FSWR-2025-0006 (НИЛ «Лаборатория мемристорной наноэлектроники»), 2025-2027 гг.); «Нейроэлектроника – интеллектуальные нейроморфные и нейрогибридные системы на основе новой электронной компонентной базы (этап 2023-2025)» (Договор № 1770641334823000080 0/96-2023/213 от 15.08.2023 г. с ФГУП «РФЯЦ-ВНИИЭФ» в рамках Научной программы Национального центра физики и математики (направление «Искусственный интеллект и большие данные в технических, промышленных, природных и социальных системах»), 2023-2025 гг.).

Основные результаты диссертационного исследования докладывались и обсуждались на 6-ти конференциях в том числе: Международная научно-техническая конференция «Информационные системы и технологии» (Нижний Новгород, НГТУ им. Алексеева; Всероссийская научная конференция «Нейрокомпьютеры и их применение» (Москва, МГППУ); Школа-конференция с международным участием «Нейроэлектроника и нейротехнологии будущего» (Нижний Новгород, ННГУ им. Н.И. Лобачевского); Scientific School Dynamics of Complex Networks and their Applications (Калининград, БФУ им. Канта); The International Conference on Industrial Engineering (Сочи, 2023); Международная конференция «Инжиниринг & Телекоммуникации» (Москва, МФТИ, 2021).

Основные теоретические и практические результаты диссертационной работы опубликованы в 14 научных трудах, из них по теме диссертационной работы 14, среди которых 4 публикаций в ведущих рецензируемых изданиях из перечня рекомендованных ВАК Минобрнауки РФ, 4 публикации, индексируемых в международной базе данных Scopus/Web of Science. Имеется 3 свидетельства о государственной регистрации программ для ЭВМ (приложение Б).

Научные статьи, опубликованные в журналах из перечня ВАК

Борданов, И. А. Оценка точности работы искусственных нейронных сетей на базе мемристивных устройств на основе теории планирования эксперимента / И. А. Борданов, Л. Я. Королёв, С. А. Щаников, А. Н. Михайлов // Радиотехнические и телекоммуникационные системы. – 2025. – № 2. – С. 40–52. – Текст : непосредственный.

Борданов, И. А. Оценка точности работы искусственных нейронных сетей на базе мемристоров с применением моделей на основе данных / И. А. Борданов, С. А. Щаников // Радиотехнические и телекоммуникационные системы. – 2024. – № 2 (54). – С. 59–68. – Текст : непосредственный.

Данилин, С. Н. Количественное определение отказоустойчивости искусственных нейронных сетей на базе мемристоров / С. Н. Данилин, С. А. Щаников, **И. А. Борданов**, А. Д. Зуев // Нейрокомпьютеры: разработка, применение. – 2020. – Т. 22, № 1. – С. 55–65. – Текст : непосредственный.

Борданов, И. А. Современное состояние в области аппаратной реализации искусственных нейронных сетей на базе мемристоров / **И. А. Борданов**, С. А. Щаников, С. Н. Данилин // Телекоммуникации. – 2020. – № 8. – С. 35–48. – Текст : непосредственный.

Научные публикации, индексируемые в международных базах Scopus и/или Web of Science

Bordanov, I. A. Determining the fault tolerance of memristorsbased neural network using simulation and design of experiments / I. A. Bordanov [et al.] // 2018 Engineering and telecommunication (EnT-MIPT). – IEEE, 2018. – P. 205-209. – Текст : непосредственный. – DOI: 10.1109/EnT-MIPT.2018.00053.

Bordanov, I. A. High-performance software for memristor-based neural network simulation and optimization / I. A. Bordanov, R. A. Mineev, S. N. Danilin. – Текст : непосредственный // 2021 International Conference Engineering and Telecommunication (En&T) / Moscow Institute of Physics and Technology – IEEE, 2021. – P. 1–4.

Bordanov, I. A. Modeling and hardware implementation of vector-matrix multiplier based on 32x8 1T1R memristive crossbar array / I. A. Bordanov [et al.] // 2023 7th Scientific School Dynamics of Complex Networks and their Applications (DCNA). – IEEE, 2023. – P. 249–251. – Текст : непосредственный. – DOI: 10.1109/DCNA59899.2023.10290511.

Bordanov, I. A. Simulation of calculation errors in memristive crossbars for artificial neural networks / I. A. Bordanov, A. A. Antonov, L. Ya. Korolev // 2023 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). – IEEE, 2023. – P. 1008–1012. – Текст : непосредственный. – DOI: 10.1109/ICIEAM57311.2023.10139308.

Тезисы докладов конференций и публикации в прочих изданиях

Борданов, И. А. Исследование влияния погрешностей матрично-векторного умножения на точность работы искусственных нейронных сетей на базе мемристоров / И. А. Борданов, С. Н. Данилин – Текст : непосредственный // Нейрокомпьютеры и их применение : сборник тезисов XXI Всероссийской научной

конференции / Московский государственный психолого-педагогический университет. – Москва, 2023. – С. 156–157.

Борданов, И. А. Применение методологии имитационного моделирования для оценки точности работы искусственных нейронных сетей на базе мемристивных устройств / И. А. Борданов, С. А. Щаников – Текст : непосредственный // Труды первой школы-конференции с международным участием «Нейроэлектроника и нейротехнологии будущего» / ННГУ. – Нижний Новгород, 2024. – С. 32.

Борданов, И. А. Оценка точности работы искусственных нейронных сетей на базе мемристоров с применением моделей на основе данных / И. А. Борданов, С. А. Щаников. – Текст : непосредственный // Информационные системы и технологии - 2025 : программа и аннотации докладов XXXI Международной научно-технической конференции / НГТУ им. Р. Е. Алексеева. – Нижний Новгород, 2025. – С. 36.

Свидетельства о государственной регистрации программы для ЭВМ

Свидетельство № 2023666086 Российская Федерация. Программа для моделирования матрично-векторного умножения с учетом погрешностей мемристивных устройств : № 2023665065 : заявлено 17.07.2023 : опубликовано 26.07.2023 / **Борданов И. А.**, Щаников С. А. – 1 с. – Текст : непосредственный.

Свидетельство № 2024619751 Российская Федерация. Программа для оценки точности работы искусственных нейронных сетей на базе мемристоров с учетом погрешности матрично-векторного умножения : № 2024616902 : заявлено 02.04.2024 : опубликовано 25.04.2024 / **Борданов И. А.**, Щаников С. А. – 1 с. – Текст : непосредственный.

Свидетельство № 2019661246 Российская Федерация. Модуль определения точности функционирования искусственных нейронных сетей на базе мемристоров для системы имитационного моделирования : № 2019619978 : заявлено 12.08.2019 : опубликовано 23.08.2019 / Щаников С. А., **Борданов И. А.**, Данилин С. А., Зуев А. Д. – 1 с. – Текст : непосредственный.

Личный вклад

Основные результаты, представленные в диссертационной работе, получены автором лично. Постановка цели и задач, обсуждение планов исследований и полученных результатов выполнены совместно с научным руководителем. Во всех совместных публикациях соискатель принимал активное участие и внес существенный вклад на всех этапах работы начиная от постановки задачи и заканчивая обработкой экспериментальных данных и формированием выводов по результатам исследования. Алгоритмы создания моделей, анализа и синтеза параметров сигналов и компьютерного моделирования ИНСМ разработаны автором лично.

Структура и объем диссертации

Диссертационная работа состоит из введения, 4 глав с выводами и заключения. Диссертационная работа изложена на 120 страницах машинописного текста и содержит 40 рисунок, 18 таблиц, 2 приложений общим объемом 6 страниц. Список литературы состоит из 115 источника.

1 Современные методы оценки и обеспечения функциональной корректности искусственных нейронных сетей на базе мемристоров

1.1 Особенности аппаратной реализации искусственных нейронных сетей

В настоящее время алгоритмы работы ИНС в основном выполняются на компьютерах с архитектурой Джона фон Неймана (рисунок 1.1). Для ускорения вычислительных операций применяются графические процессоры, которые хорошо подходят для данной задачи благодаря высокому параллелизму вычислений, что в некоторой степени соответствует принципам функционирования биологических нейронных сетей [4,5]. Одним из основных минусов данной реализации является высокое энергопотребление [6]. Это связано с тем, что в архитектуре Джона фон Неймана память физически отделена от вычислительного устройства и большое количество вычислительных мощностей тратится на пересылку данных между устройством, отвечающим за хранение параметров ИНС и промежуточных результатов, и устройством, выполняющим вычисления [7].

Рисунок 1.1 – Варианты аппаратной реализации ИНС

В связи с интенсивным развитием технологий ИИ повышается общая энергетическая нагрузка на электростанции, что с одной стороны вызывает дефицит электроэнергии, а с другой стороны может привести к негативным последствиям для окружающей среды [8]. Говоря о мобильных бортовых и носимых системах, высокое энергопотребление усложняет задачу реализации сложных вычислений на устройстве, повышения их конфиденциальности,

безопасности и т.д. [9] Поэтому в настоящие время многие научные коллективы из различных стран и крупных технологических компаний, таких как IBM, Intel и др., ищут возможности повышения энергоэффективности вычислителей для приближения их к энергоэффективности мозга [10,11].

Одним из возможных решений является разработка нейроморфных процессоров. Это специализированные вычислительные устройства, архитектура которых оптимизирована для максимально эффективного выполнения алгоритмов работы ИНС и нейроморфных систем. Под нейроморфными системами понимается такой класс систем, алгоритмы работы которых базируются на принципах функционирования работы головного мозга. В данном направлении исследований можно выделить два основных подхода (рисунок 1.1):

– Вычисления рядом с памятью.

– Вычисления в памяти.

При вычислениях рядом с памятью каждое процессорное ядро имеет свою локальную память, в которой хранятся значения весов синапсов ИНС. Одним из первых таких устройств был исследовательский нейрочип TrueNorth, разработанный сотрудниками компании IBM [12]. В последствии были разработаны и другие нейрочипы, основанные на подобных принципах, такие как Loihi от американской компании Intel [11], «Алтай» от российской компании «Мотив Нейроморфные Технологии» [13], Tianjic, разработанный учеными из университета Цинхуа [14] и многих других. Данные нейроморфные процессоры являются полностью цифровыми, в качестве памяти для синапсов обычно используется SRAM.

Все описанные выше нейрочипы имеют лучшую энергоэффективность по сравнению с графическими и тензорными процессорами. Например, пиковая производительность нейрочипа Tianjic достигает 1,28 TOPS при частоте 300 МГц, что при решении задачи распознавания рукописных цифр на наборе данных MNIST позволяет обеспечить в 2-3 раза более высокую энергоэффективность по сравнению с графическим процессором NVIDIA Titan-Xp.

Концепция вычислений непосредственно в памяти заимствована из биологии – в мозге нет отдельно процессора и устройства хранения, синапсы хранят веса и преобразуют сигналы. Такой подход позволяет преодолеть проблему бутылочного горлышка архитектуры Джона фон Неймана, в связи с чем концепция вычислений в памяти потенциально позволяет создавать еще более энергоэффективные и производительные вычислители. Важную роль в разработке таких устройств играет выбор электронных компонентов для организации памяти, чьи характеристики напрямую влияют на энергоэффективность и производительность нейроморфных систем.

В настоящее время существует несколько основных компонентов которые могут быть использованы для реализации концепции вычислений в памяти, а именно технологии памяти на основе заряда (SRAM, DRAM, FLASH) и на основе сопротивлений (STT- MRAM, PCM, ReRAM).

SRAM ячейки при реализации нейроморфных систем в основном используются для реализации операции умножения и накопления непосредственно в памяти. Такое применение SRAM реализовано в работе [15]. В других подобных работах также предлагаются свои варианты использования SRAM для реализации операции умножения и накопления, такие как XNOR-SRAM с восемью транзисторами [16] или TD-SRAM [17], в которой совмещают парадигму вычислений в памяти с парадигмой вычислений во временной области.

Еще одной распространённой технологией памяти на основе заряда является DRAM, которая также применяется для аппаратной реализации нейроморфных систем. В работе [18] авторы используя память DRAM, а именно её мобильного варианта LPDDR4, реализовали цифровой ускоритель искусственного интеллекта McDRAM v2, который имеет в 1,7 раза лучшую производительность, в 3,7 раза лучшее энергопотребление и в 8,6 раз лучшую масштабируемость по сравнению с мобильным ускорителем Jetson AGX Xavier и в 9,3 раза выше энергоэффективность по сравнению с ускорителем серверного класса NVIDIA TITAN RTX.

NAND это тип FLASH памяти, который имеет высокую плотность, может хранить большие значения весов синапсов и имеет способность к выполнению

матрично-векторного умножения в аналоговом виде. В работе [19] демонстрируются возможности применения NAND для реализации ИНС при этом прогнозируемая энергоэффективность данного устройства составляет 1,15–19,01 TOPS/Вт.

Память на основе фазового перехода является одним из видов резистивной памяти, которая также используется для реализации нейроморфных систем. Возможности данной памяти наглядно были продемонстрированы в работе [20]. Авторы представили 64 ядерный исследовательский чип IBM HERMES Project Chip, производительность которого составляет 63,1 TOPS, а энергоэффективность 9,76 TOPS/Вт для умножения 8-битной матрицы весов на вектор.

Еще одним типом резистивной памяти является магниторезистивная память с произвольным доступом (MRAM). В работах [21–23] демонстрируются возможности применения STT-MRAM памяти, реализованной в виде структур 1T1J и 2T2J, для создания бинарных нейронных сетей различных архитектур, таких как сверточная нейронная сеть, спайковая нейронная сеть и многослойный персептрон.

Одним из наиболее перспективных типов памяти для реализации нейроморфных систем является ReRAM или RRAM, которая реализуется на базе мемристивных устройства. Одними из основных преимуществ данных устройств являются возможность энергонезависимого хранения весов синапсов нейронов в виде сопротивлений в отличии от SRAM и DRAM. Кроме того, одно мемристивное устройство может иметь различные резистивные состояния, к примеру в работе [24] удалось достигнуть 2048 стабильных состояний что соответствует 11-битному разрешению ячейки памяти.

Аппаратная реализация нейроморфных систем на базе мемристивных устройств демонстрируется во многих работах [25–45]. При этом разработанные нейроморфные системы на их основе достигают отличных показателей энергоэффективности по сравнению с другими видами памяти. К примеру, в работе [30] представлено устройство для аппаратной реализации сверточных нейронных сетей. Данное устройство показывает примерно в 110 раз лучшую

энергоэффективность и в 30 раз лучшую производительность по сравнению с NVIDIA Tesla V100.

Несмотря на явные преимущества, мемристивным устройствам свойственны вариации сопротивлений как от цикла к циклу записи, так и от устройства к устройству. В результате этого достаточно сложно установить сопротивление мемристивного устройства с абсолютной точностью на требуемое значение, без использования окна допуска. При этом, даже если удастся установить нужное сопротивление, то оно все равно будет варьироваться в определенном диапазоне в процессе функционирования. Основным вкладом в такие флуктуации является случайный телеграфный шум, который характеризуется ступенчатыми переходами между двумя или более уровнями тока в произвольные моменты времени при постоянном напряжении считывания [24].

Вариации сопротивлений мемристивных устройств вызывают вариации и весовых коэффициентов синапсов нейронов ИНС, что приводит к появлению погрешностей вычислений матрично-векторных умножений и, как следствие, к изменению значений метрик ФК ИНСМ [1,2].

1.2 Факторы, влияющие на функциональную корректность искусственных нейронных сетей на базе мемристоров

Несмотря на хорошие перспективы аппаратной реализации ИНС на базе мемристивных устройств, существует определенная проблема обеспечения требуемого значения метрик ФК, которая в значительной степени связана с ограниченной точностью задания сопротивлений мемристивных устройств и возможными вариациями сопротивлений в процессе функционирования. В общем случае можно выделить следующие две основные группы факторов, разделенные по типу оказываемого ими влияния на ФК ИНСМ [46]:

– Факторы, оказывающие косвенное воздействие.

– Факторы, оказывающие прямое воздействие.

К первой группе можно отнести: схему интеграции мемристоров, вариации от устройства к устройству, стойкость к переключению, тип устройства, параметры окружающей среды, время прошедшее с последнего момента установки сопротивления мемристивного устройства, состояние сопротивления устройства.

Ко второй группе можно отнести: минимальное и максимальное сопротивление, наименьшее воспроизводимое изменение сопротивления, нелинейность изменения сопротивления, асимметрия изменения сопротивления, вариация сопротивления от цикла к циклу, нелинейность и асимметрия ВАХ, дрейф состояний сопротивлений, случайный телеграфный шум.

На рисунке 1.2 показана взаимосвязь факторов косвенного и прямого воздействия, а также их влияние на ФК ИНСМ.

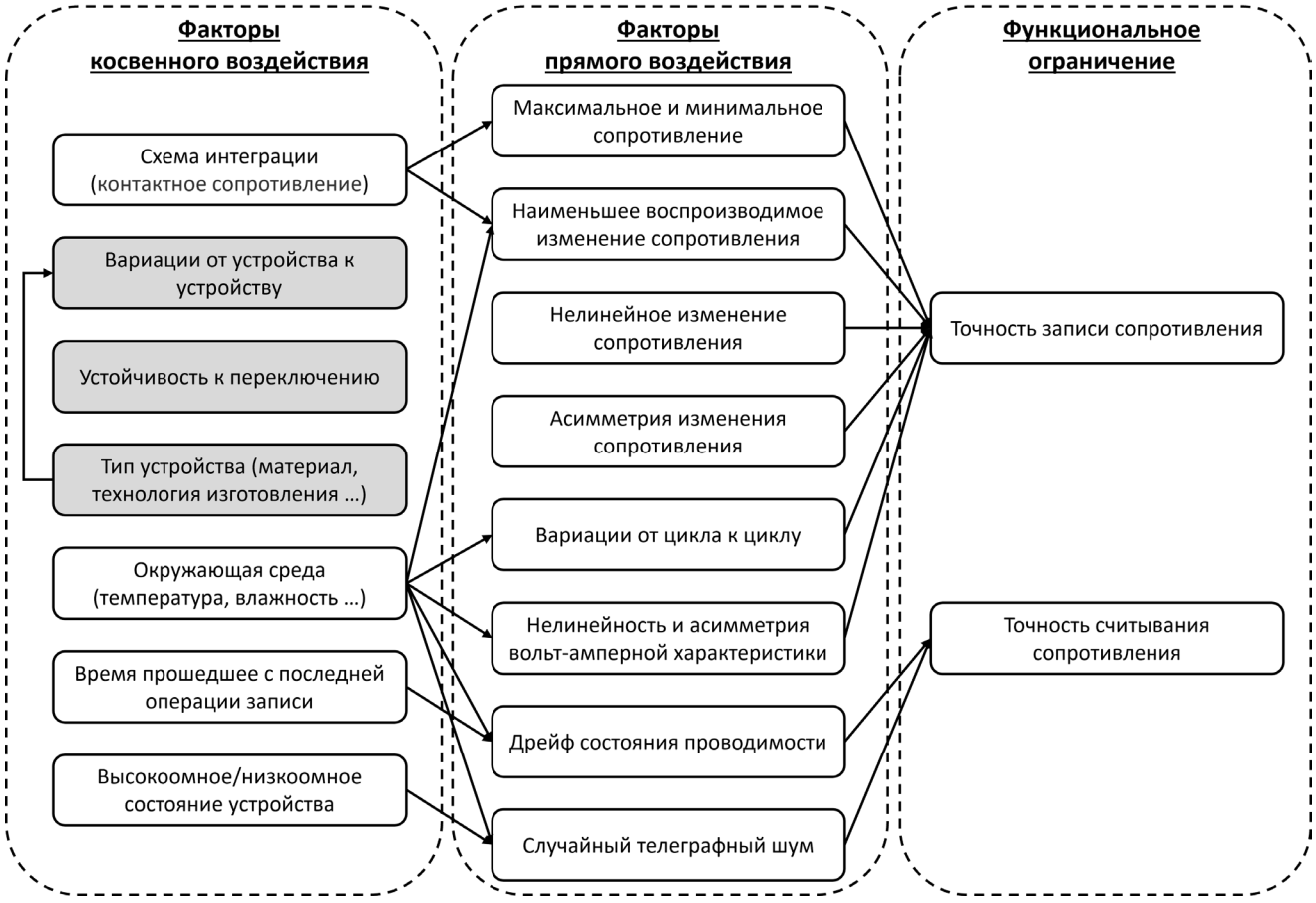


Рисунок 1.2 – Влияние параметров мемристивных устройств на ФК ИНСМ. Серые прямоугольники влияют на все параметры прямого воздействия

Из данного рисунка видно, что все факторы влияют на 2 основных показателя – точность записи сопротивления и точность считывания сопротивления. Из данных факторов особенно стоит выделить вариации от устройства к устройству,

нелинейность изменения сопротивления, вариации от цикла к циклу, дрейф состояний сопротивлений и случайный телеграфный шум. Именно они оказывают значительное влияние на вариации сопротивлений мемристивных устройств и соответственно на ФК ИНСМ.

При моделировании ИНСМ необходимо учитывать данные факторы, а также степень их влияния на функционирование мемристивных устройств в составе ИНСМ. Это является важным для оценки и обеспечения ФК нейроморфных систем на этапе проектирования.

1.3 Методы обеспечения функциональной корректности искусственных нейронных сетей на базе мемристоров

В соответствии с ГОСТ 59898-2021 метрики функциональной корректности (правильности) используются для оценки обеспечения степени точности результатов, а также частоты встречаемости ошибок и недопустимых отклонений в системах ИИ. Выбор метрики зависит от типа решаемой задачи, а именно для задач регрессии используется средняя квадратичная ошибка и средняя абсолютная ошибка, для задач классификации и обнаружения – доля правильных исходов (accuracy), точность (precision), чувствительность (sensitivity), избирательность (specificity), *F*-мера, площадь под кривой ROC и площадь под кривой PRC, для задач ранжирования – приведенная суммарная эффективность, для задач восстановления (синтеза и реконструкции) изображений – пиковое отношение сигнал/шум и индекс структурного сходства.

Говоря об ИНСМ, значения данных метрик можно разделить на 3 вида – достигнутые на программной модели ИНС в результате обучения, оцененные при моделировании ИНСМ и рассчитанные в результате функционирования в аппаратном варианте исполнения. Если ИНС обучена корректно, то наличие вариаций сопротивлений мемристивных устройств и весов синапсов обычно вызывает ухудшение значений метрик ФК второго и третьего вида. Если ИНС

недообучена, то может наблюдаться и улучшение значений метрик ФК, но данный случай скорее является исключением. Поэтому задачу обеспечения ФК ИНСМ можно сформулировать как задачу минимизации разности между достигнутым в результате обучения программной модели значением метрики ФК и значением, полученным при аппаратной реализации. Данная задача может быть решена различными способами, показанными на рисунке 1.3.

Рисунок 1.3 – Методы обеспечения ФК ИНСМ

1.3.1 Совершенствование материалов и структур мемристивных устройств

Ключевым фактором обеспечения ФК ИНСМ является создание мемристивных устройств с высокой стабильностью электро-физических характеристик. Данная задача решается через поиск новых материалов, обладающих мемристивными свойствами, а также разработкой новых вариантов мемристивных структур [46].

Мемристивные устройства могут создаваться как из органических, так и из неорганических материалов, каждый из которых имеет свое целевое применение, преимущества и недостатки. В области органических материалов значительных

успехов достиг научный коллектив из НИЦ «Курчатовский институт». В статьях [47,48] описан вариант органического мемристора из полианилина, обладающий высокой стабильностью электро-физических характеристик для данного класса устройств.

Однако большинство мемристоров реализуется на основе неорганических материалов. Например, в статье [49] авторы для создания мемристора использовали структуру TiN/TaO_x/$HfAl_yO_x$/TiN. Особенностью данного мемристора является наличие слоя TaO_x толщиной 60 нм. Данный слой работает как ограничитель тока и помогает модулировать сопротивление мемристивного устройства, что впоследствии позволяет получить более высокое значение сопротивления. Для создания мемристоров также применяют структуры Al/GO/n-Si, которые демонстрируют безформовочные биполярные резистивные характеристики переключения с плотностью тока до 1 А/см^2 [50], Ti/ZnO/Pt, проявляющие стабильный мемристивный эффект, мало зависящий от морфологии нанокристалических пленок оксида цинка [51], Pt/TiO_2/Al_2O_3/Pt [52], демонстрирующее стабильное биполярное переключение относительно резистивного состояния, определяемого напряжением смещения, Металл/$(Co_{41}Fe_{39}B_{20})_x(LiNbO_3)_{100-x}$/Металл [53], с отношением его сопротивления между высоким и низким состоянием равным примерно 65, Si/SiO_2/Si [54], имеющим высокий коэффициентом проводимости (10^4) и длительное хранение данных при повышенной температуре 300 °C и многие другие структуры [55–57,57–61] со своими уникальными преимуществами и характеристиками.

Анализ данных публикаций позволяет сделать вывод, что существующие технологии всё еще далеки от совершенства с точки зрения повторяемости ВАХ от цикла к циклу, но что еще более важно – от устройства к устройству. Однако, несмотря на это, данные вариации носят не случайный характер, а могут быть описаны статистически и позволяют уже сейчас разрабатывать варианты аппаратной реализации ИНСМ и проводить их исследования.

1.3.2 Оптимизация настройки искусственных нейронных сетей на базе мемристоров

Настройка ИНСМ заключается в задании определённых весов синапсов нейронов путем изменения сопротивлений мемристивных устройств. Значения весов могут быть получены в результате обучения ИНСМ разными способами, а именно с применением программной модели, с использованием различных аппаратных схем обучения и в гибридном формате (программно-аппаратное обучение).

Один из наиболее распространенных подходов к обеспечению требуемой ФК заключается в программном обучении модели ИНСМ в одном из фреймворков, таких как TensorFlow, PyTorch и др. После получения весов, они пересчитываются в сопротивления мемристивных устройств, с учетом особенностей применяемых схемотехнических решений, и выполняется их маппирование. В частном случае, в процессе программного обучения могут учитываться вариации сопротивлений мемристоров, задаваемые как погрешности к значениям весов, получаемым от эпохи к эпохе. Например, в работе [62] авторами представлен алгоритм программного обучения ИНСМ с учетом неидеальностей мемристивных устройств. В результате моделирования сети для классификации изображений из датасета MNIST было выявлено, что учет данных погрешностей позволяет понизить ошибку сети до 9,1% при этом регуляризация позволяет снизить ошибку еще ниже до 7,1%. Подобный подход применялся и в работе [63] благодаря чему удалось повысить долю правильных исходов ИНСМ на 0,02.

Обучение ИНСМ может осуществляться непосредственно в нейрочипе, что позволяет в процессе обучения учитывать все погрешности используемых мемристивных устройств. В статье [64] описан подход к обучению клеточной нейронной сети с применением алгоритма случайного изменения веса, реализованного аппаратно. Основным недостатком алгоритма является большая

продолжительность процесса обучения ИНСМ. В работе [65] авторы применили метод аппаратного обучения спайковой нейронной сети с двумя слоями нейронов.

Выполнение обучения ИНСМ на реальных устройствах делает ИНСМ более устойчивыми к неидеальностям, таким как наличие неисправных устройств и вариаций от устройства к устройству [66]. Однако, при аппаратном обучении существует значительная вероятность того, что сеть вообще не достигнет требуемой ФК в процессе обучения для заданной архитектуры. Кроме того, аппаратное обучение не учитывает флуктуации сопротивлений, возникающие в процессе работы ИНСМ.

Существуют также гибридные алгоритмы обучения ИНСМ. В работе [67] предложен алгоритм в котором ИНСМ сначала обучается программно, затем веса переносятся в устройство и сеть дообучается аппаратно. В результате было доказано, что такой подход позволяет сократить время обучения по сравнению с чисто аппаратным, а также снизить ошибку программного подхода при переносе весов на аппаратную базу. Другим вариантом гибридного подхода является многократный перенос весов в реальное устройство в процессе обучения программной модели между эпохами обучения. В данном случае программная модель возвращает целевые поправки к весам, которые записываются в устройства, затем считываются реальные значения и возвращаются в программную модель. Это позволяет повысить точность обучения, но делает его более долгим. Таким образом, гибридный подход имеет ряд преимуществ над предыдущими подходами, однако предполагает двойную сложность.

Существует множество различных методов настройки нейроморфных систем на базе мемристивных устройств, однако универсального подхода не существует.

1.3.3 Оптимизация структуры и архитектуры искусственных нейронных сетей на базе мемристоров

Основной вариант обеспечения ФК ИНСМ с точки зрения архитектуры заключается в применении мажоритарного элемента на выходе ансамбля из нескольких моделей. Пример такого подхода описан в работе [68]. В результате исследований авторы показали, что применение комитета машин позволяет приблизить реальное значение используемой метрики ФК ИНСМ к программному за счет увеличения количества моделей.

К структурным подходам можно отнести способ организации синаптических связей в нейроморфных системах, а именно определение того, какое количество мемристоров использовать для их организации. Например, в статьях [69,70] описан подход, при котором каждый синапс ИНСМ реализован на базе двух мемристоров. Такая архитектура делает ИНСМ более устойчивой к дефектным мемристорам, «залипшим» в состоянии низкого или высокого сопротивления, а также позволяет уменьшить потребление энергии практически в 9 раз, по сравнению с подходом, который использует только один мемристор на синапс [71].

Существуют примеры применения большего числа мемристоров в составе синапса. Например, в статье [72] каждый синапс ИНСМ реализуется на базе четырех мемристоров, а в работе [73] каждый синапс нейронной сети реализован на базе девяти мемристоров. Такая архитектура в совокупности с предложенной авторами схемой восстановления сопротивления может уменьшить процентное изменение сопротивления с 21,1% до 0,12% в процессе работы ИНСМ. Синапсы ИНСМ также могут быть реализованы на базе одного мемристора. Такие архитектуры представлены в работах [74].

Подводя итог по методам обеспечения ФК ИНСМ, описанным в пункте 1.3, можно сделать вывод о том, что не существует единственного идеального подхода к решению данной задачи. Описанные в данном пункте подходы позволяют обеспечить ФК и надежность работы ИНСМ на определенном уровне, при этом

некоторые подходы можно комбинировать друг с другом, улучшая итоговый вариант. Для того, чтобы на этапе проектирования нейроморфной системы можно было понять, как будут влиять на ФК ИНСМ используемые мемристоры, методы настройки весов, параметры структуры и архитектуры системы, необходимо уметь оценивать ФК на моделях.

1.4 Методы оценки функциональной корректности искусственных нейронных сетей на базе мемристоров

1.4.1 Основные подходы к оценке функциональной корректности искусственных нейронных сетей на базе мемристоров

Основные подходы к оценке ФК ИНСМ представлены на рисунке 1.4. Для простых вариантов ИНСМ, состоящих из небольшого количества устройств, ФК можно попытаться рассчитать аналитически. Для этого необходимо иметь формулы, описывающие ВАХ мемристоров, формулы вариаций сопротивлений, а также все формулы, описывающие функционирование всей схемотехники ИНСМ. Основное ограничение данного подхода заключается в том, что для каждого мемристивного устройства приходится индивидуально выполнять сложный процесс идентификации, что требует значительных ресурсов для нейрочипов, содержащих миллионы мемристивных устройств.

Рисунок 1.4 – Основные подходы к оценке ФК ИНСМ

Другим вариантом оценки ФК ИНСМ является разработка, конструирование и изготовление материального прототипа (опытного образца). У такого подхода есть одно основное преимущество – результат оценки ФК ИНСМ на основе материального прототипа будет максимально приближен к итоговому изделию. Однако у данного подхода есть ряд недостатков, а именно высокие риски недостижения требуемого значения метрики ФК ИНСМ, временные и финансовые затраты/потери в случае неудачи, необходимость задействования профильных специалистов (конструктора/монтажники/производители плат). Таким образом, использование данного метода более предпочтительно на финальной стадии проектирования, когда все риски уже оценены и применены соответствующие меры для обеспечения ФК ИНСМ.

В связи с этим на этапе проектирования ИНСМ предпочтительнее использовать подходы, основанные на компьютерном моделировании. Существует два основных вида имитационных моделей [75], которые могут быть использованы в процессе оценки ФК ИНСМ: низкоуровневые модели на уровне схем и высокоуровневые модели на уровне систем.

Низкоуровневое моделирование позволяет учитывать различные особенности аппаратной реализации ИНСМ на уровне схем (работу АЦП, ЦАП, мемристивных устройств, операционных усилителей и др.), что обеспечивает высокий уровень точности результатов моделирования. В настоящее время существует большое количество готовых моделей мемристивных устройств [75] для программ схемотехнического моделирования.

Основной недостаток схемотехнического моделирования схож с недостатком аналитического подхода – несмотря на то, что с помощью компьютера можно быстрее рассчитывать параметры системы в соответствии с математическими моделями, параметры данных моделей требуют настройки для каждого мемристора индивидуально, а также требуют настройки параметров модели вариаций сопротивлений. Кроме того, для устройств содержащих миллионы элементов, компьютерный расчёт требует высокой вычислительной мощности и может приводить к значительным временным затратам, которые увеличиваются в

зависимости от масштаба модели [76]. Также при таком подходе необходимо обладать знаниями низкоуровневого моделирования схем и инструментов, которые во многом являются не стандартизированными [75].

Применение высокоуровневого моделирования подразумевает описание модели ИНСМ на языке высокого уровня без моделирования конкретных факторов, влияющих на ФК, но с учетом степени их общего влияния на верхнеуровневые компоненты системы. Например, погрешность сопротивления вызывает погрешность веса синапса нейрона, соответственно при высокоуровневом моделировании ИНСМ нет необходимости моделировать отдельное мемристивное устройство, достаточно смоделировать вес. В большинстве публикаций авторы моделируют вес статистически, задавая его значения по определенному закону распределения (нормальному, Релея и др.) с заданным фиксированным диапазоном отклонений. Данный подход позволяет относительно быстро оценить эффективность выбранной архитектуры, структуры или методов борьбы с погрешностями мемристивных устройств для обеспечения ФК ИНСМ [75,76].

Основным недостатком высокоуровнего моделирования является более низкая точность результатов оценки, поскольку при моделировании пренебрегают детальной информацией о параметрах отдельных компонентов ИНСМ. Кроме того, в существующих моделях параметры разброса сопротивлений в основном оценочно извлекаются из вольтамперных характеристик мемристоров или характеристик синаптической пластичности. При таком подходе невозможно установить взаимосвязь между весом и параметрами сигнала записи сопротивления, поскольку сигнал снятия вольтамперной или другой характеристики является лишь частным случаем сигнала, с помощью которого может быть задано конкретное сопротивление мемристора.

Таким образом, все подходы к оценке ФК ИНСМ важны – высокоуровневое имитационное моделирование позволяет оценить ФК ИНСМ на ранних этапах проектирования и, в случае неудовлетворительных результатов, принять соответствующие меры по обеспечению ФК. Далее, отдельные компоненты

системы можно смоделировать на уровне схем с учетом схемотехнических особенностей и создать прототип. Такая последовательность процесса проектирования ИНСМ позволяет уменьшить риски при разработке.

1.4.2 Основные виды моделей мемристивных устройств

В общем случае подходы к построению моделей мемристивных устройств можно разделить на три группы (рисунок 1.5):

– Модели, которые строятся на основе знаний о физических свойствах материалов и структур мемристивных устройств (физические модели).

– Модели, которые строятся на основе экспериментальных данных, без знания физических свойств материалов и структур мемристивных устройств (эмпирические модели).

– Модели, которые строятся и на основе знаний о физических свойствах материалов и структур и на основе экспериментальных данных (полуэмпирические модели).

Модели, которые строятся на основе знаний о физических свойствах материалов и структур мемристивных устройств (физические модели), могут адекватно описывать работу мемристивных устройств и хорошо подходят для изучения физических процессов, происходящих в них [77]. Однако они имеют высокую вычислительную сложность, что не позволяет в полной мере использовать их при моделировании влияния флуктуаций мемристивных устройств на ФК ИНСМ [3].

Рисунок 1.5 – Виды моделей мемристивных устройств

Полуэмпирические или физически обоснованные компактные модели являются упрощенным вариантом физических моделей с допущением об идеальности некоторых параметров структуры устройства [78]. Это позволяет

снизить вычислительные затраты на эмуляцию таких моделей и открывает широкие возможности для применения их при моделировании ИНСМ на низком уровне в схемотехнических симуляторах.

В настоящее время одной из наиболее распространенных и относительно простых компактных моделей является Стэнфордская модель. Например, в работе [3] авторы использовали данную модель для моделирования вариаций от цикла к циклу. При этом важно отметить, что для подгонки модели к экспериментальным данным, полученным с реального мемристивного устройства, требуется указать много различных параметров, таких как толщину диэлектрика, расстояние между атомами, коэффициент усиления, учитывающий поляризуемость материала и т.д. Значения данных параметров напрямую влияют на адекватность моделей и их идентификация требует не только знания характеристик материалов и структур, но и проведения ряда экспериментов для определения физических свойств [79–84].

Эмпирические модели строятся на основе экспериментальных данных, получаемых с мемристивных устройств. В частности, такие модели применяются для оценки вариаций мемристивных устройств, что наглядно демонстрируется в работе [3]. В ней авторы с помощью временных рядов смоделировали вариации напряжения сброса и установки сопротивления в зависимости от цикла к циклу, при этом временные ряды позволяют учитывать корреляцию между циклами. Подобный подход к моделированию мемристивных устройств, но со своими особенностями, представлен в работе [85]. В ней авторы демонстрируют разработанный критерий для оценки вариаций мемристивного устройства от цикла к циклу, который не только учитывает напряжение установки и сброса сопротивления, но и форму ВАХ.

Анализ публикаций показал, что модели, не учитывающие физические свойства материалов, с одной стороны хорошо подходят для моделирования вариаций мемристивных устройств [86–92], с другой стороны просты в создании, поскольку не требуют детального исследования физических процессов, происходящих в мемристоре, поэтому данный подход к созданию моделей может применяться и для оценки ФК ИНСМ на верхнем уровне.

1.4.3 Программное обеспечение для моделирования искусственных нейронных сетей на базе мемристоров

Программы для моделирования ИНСМ можно разделить на две основных группы (рисунок 1.6), а именно программы, предназначенные для моделирования процесса работы ИНСМ на конкретных аппаратных ускорителях ИИ и универсальные.

Программы, предназначенные для моделирования процесса работы ИНСМ на конкретных аппаратных ускорителях ИИ, позволяют продемонстрировать то, как ИНС будет преобразовывать данные при аппаратной реализации на данном ускорителе. К таким программам можно отнести пакеты RAPIDNN [93] или PUMA [94], которые являются эмуляторами одноименных ускорителей искусственного интеллекта. Несомненным преимуществом данных программ является высокая степень соответствия архитектуре и структуре соответствующего ускорителя ИИ, для которого они предназначены. Однако они не подходят для моделирования работы ИНСМ с иной архитектурой и структурой.

Рисунок 1.6 – Виды ПО для имитационного моделирования ИНСМ

К универсальным программам можно отнести библиотеку MemTorch [75], которая может применяться для моделирования глубоких ИНСМ для полносвязных и сверточных архитектур. MemTorch напрямую интегрируется с библиотекой PyTorch и позволяет преобразовывать обученные там сети в ИНСМ. Еще одним примером является программный инструмент, описанный в статье [95], для моделирования работы нейроморфных архитектур, в частности, спайковых нейронных сетей на базе мемристивных устройств. Существует также множество

других подобных инструментов, а именно MNSIM [96], DNN+NeuroSim [97], IBM Analog Hardware Acceleration Kit [98], DL-RSIM [99] со своими уникальными возможностями и особенностями.

1.5 Обоснование разработки новых моделей и алгоритмов для оценки функциональной корректности искусственных нейронных сетей на базе мемристоров

В предыдущих пунктах данной главы были рассмотрены различные виды моделей мемристоров, а также существующие программные решения для их моделирования. В процессе проведения анализа было выявлено, что:

– существует много моделей мемристоров (абстрактных (Biolek, VTEAM, Joglekar, и т.д) или эмпирических), результатом работы которых являются вольт-амперные характеристики и в настоящий момент нет единых критериев их выбора и подгонки параметров для конкретных мемристоров;

– в моделях ВАХ есть сложности и неоднозначности в подходах к учету и заданию вариаций сопротивления от цикла к циклу и от устройства к устройству;

– моделирование ИНСМ с применением моделей ВАХ – ресурсоемкий процесс, требующий решения одного или нескольких дифференциальных уравнений для каждого мемристора, что критично для больших ИНСМ состоящий из тысяч или даже миллионов МУ;

– для каждой новой комбинации материалов и структур изменяется характеристика ВАХ, поэтому для моделей ВАХ каждый раз требуется актуализация параметров или описаний физических процессов, что не реализуемо с инженерной точки зрения, когда разработчик рассматривает мемристор как один из видов электронной-компонентной базы;

– существует подход, при котором вариации сопротивлений заменяют вариациями весов ИНСМ, задавая их по одному из законов распределения с одинаковой погрешностью и пренебрегая формализацией взаимосвязи между

параметрами этого закона распределения и разбросами характеристик устройств и сигналов;

– для инженерных задач важен не только анализ ФК, но и синтез конкретных значений параметров сигналов и допусков.

Соответственно, оценка ФК ИНСМ должна рассматриваться с учетом взаимосвязи между параметрами сигналов задания сопротивления мемристивного устройства – факторами, которые влияют на итоговые сопротивления – и весами, которые влияют на метрику ФК ИНСМ. У сигнала, изменяющего сопротивление мемристора, может быть много различных параметров, число комбинаций которых очень большое.

Например, для импульсного сигнала (рисунок 1.7) в общем случае таких факторов может быть 3 – амплитуда импульса *u*, количество импульсов *n*, длительность импульса *t*. Длительность импульса связана с частотой и скважностью *s*. Возможно, что частота и скважность *s* также влияют на динамику резистивного переключения при одинаковом значении *t*, поэтому с точки зрения рассматриваемого метода они могут выступать как дополнительные факторы, увеличивающие количество возможных комбинаций сигналов. Кроме того, факторами могут являться не сами параметры импульсов, а параметры функций, которыми они описываются.

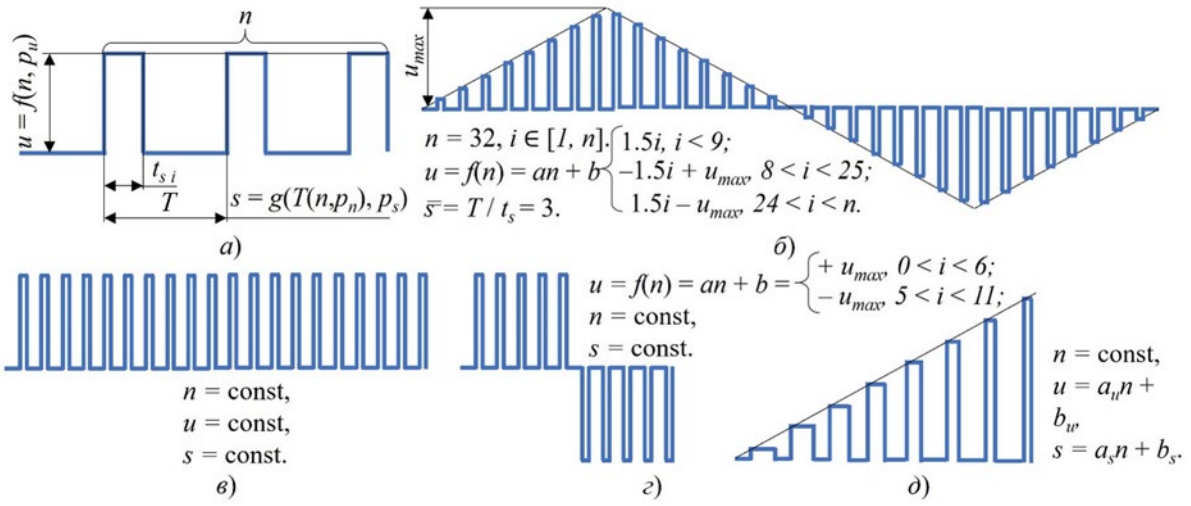


Рисунок 1.7 – Параметры импульсного сигнала, влияющие на сопротивление мемристора: а) обобщенная схема импульсного сигнала; б) пример сигнала для снятия ВАХ мемристора; в) пример сигнала для оценки времени удержания

сопротивления мемристивным устройством; г) пример сигнала для оценки количества циклов резистивного переключения мемристора; пример сигнала записи заданного резистивного состояния мемристора

Таким образом, если взять хотя бы 3 фактора с 10 уровнями, то по формуле комбинаторики размещения с повторениями количество экспериментов будет равно 1000 и это без учета того, что каждый из этих экспериментов нужно выполнить еще несколько раз, допустим 1000, что соответствуют 1000000 экспериментов. Провести все эксперименты, покрывающие все комбинации параметров сигналов невозможно. Для решения такой задачи подходит теория планирования эксперимента. В таком случае мемристор рассматривается как черный ящик (рисунок 1.8 а), а модель погрешности будь то мемристивного устройства или веса, будет определяться как некоторая функция, построенная на основе отклика системы на определенные факторы в соответствии планом эксперимента (рисунок 1.8 б).

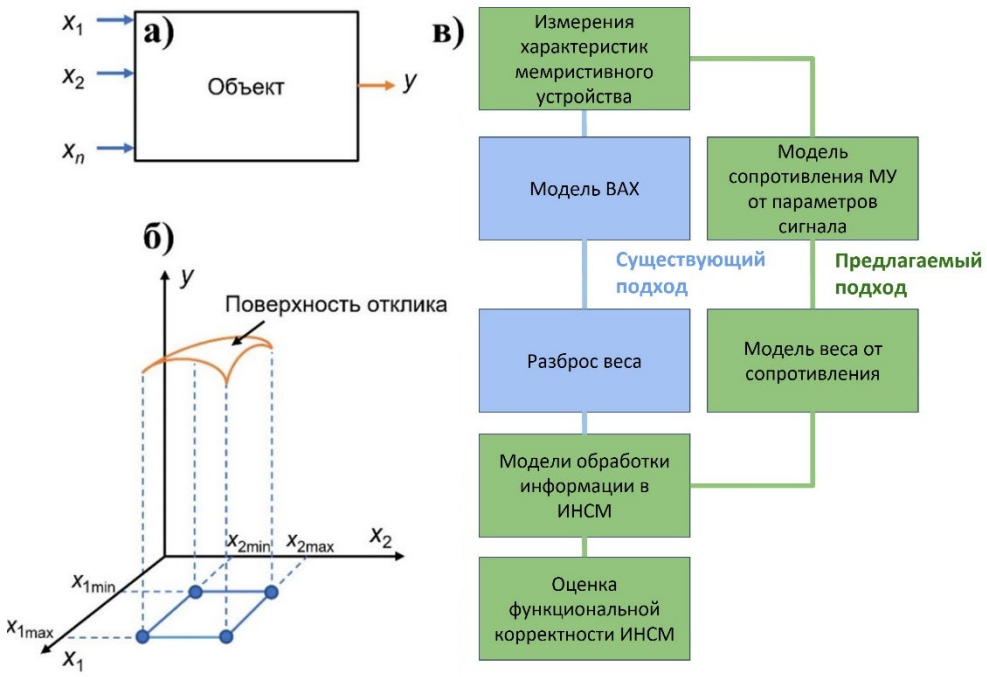


Рисунок 1.8 – Применение ТПЭ для создания моделей погрешности МУ и веса синапса нейрона: а) подход «черный ящик»; б) применение теории планирования; в) основные подходы к решению проблемы

Основываясь на анализе рассматриваемых методов и их ограничений, была выдвинута гипотеза: для повышения степени точности результатов оценки ФК

ИНСМ на этапе исследовательского проектирования необходимо учитывать взаимосвязь между параметрами сигнала задания сопротивления МУ F, результатами экспериментов по заданию сопротивлений R, значениями весов синапсов W ИНСМ и значениями метрики оценки ФК ИНСМ L (рисунок 1.9). Сопротивления R могут быть заданы только с определенной погрешностью ΔR, в результате чего возникают погрешности весов ΔW, в результате которых в свою очередь возникают изменения значений метрики оценки ФК ΔL.

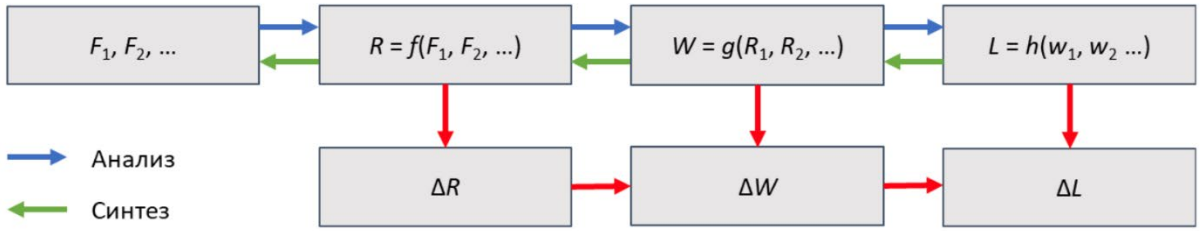


Рисунок 1.9 – Взаимосвязь между ФК и параметрами ИНСМ

1.6 Выводы по главе

1) Основным вариантом аппаратной реализации ИНС в настоящее время является их эмуляция на ЭВМ с архитектурой Джона фон Неймана. У такого подхода есть ряд недостатков, связанных, с высокими вычислительными расходами на передачу данных между памятью и вычислителем, что приводит к высокому энергопотреблению. Более перспективной является концепция вычислений в памяти, в которой вычислитель и память являются одним устройством. Для реализации данной концепции могут применяться различные электронные компоненты, однако одними из наиболее перспективных являются мемристивные устройства. Объединение таких устройств в кроссбар-массивы позволяет хранить весовые коэффициенты синапсов нейронов и выполнять матрично-векторное умножение в аналоговом виде за один такт, реализуя тем самым концепцию вычислений в памяти, что позволяет повысить скорость работы

и снизить энергопотребление ИНСМ, по сравнению с аналогами на ЭВМ с архитектурой Джона фон Неймана.

2) При разработке ИНСМ необходимо обеспечить ФК, на которую влияют ограниченная точность записи сопротивлений мемристивных устройств и их вариации в процессе работы. Данную задачу можно решить путем совершенствования материалов и структур мемристивных устройств для обеспечения высокой стабильности электрофизических характеристик, а также за счет оптимизации процедур настройки ИНСМ и подбора ее структуры и архитектуры. Поиск среди большого количества комбинаций технологических, конструкторских и алгоритмических решений вызывает необходимость в моделях и алгоритмах быстрой оценки ФК ИНСМ.

3) Для оценки ФК ИНСМ могут применяться различные подходы такие как аналитический расчет, имитационное моделирование и прототипирование. Каждый из представленных подходов может применятся на различных этапах проектирования для уменьшения возможных рисков несоответствия качества финального варианта. Также не существует универсальной модели мемристивного устройств. При моделировании ИНСМ применяются физические, эмпирические и полуэмпирические модели отличающиеся как по трудоемкости построения, так и по степени адекватности описываемому мемристивному устройству.

4) Основываясь на анализе данных особенностей и ограничений существующих подходов, в рамках диссертационного исследования были сформулированы основные пути их усовершенствования. Один из основных тезисов заключается в том, что оценка точности ИНСМ должна рассматриваться с учетом взаимосвязи между параметрами сигналов задания сопротивления мемристивного устройства – факторами, которые влияют на итоговые сопротивления – и весами, которые влияют на ФК ИНСМ. У сигнала, изменяющего сопротивление мемристора, может быть очень много различных параметров, число комбинаций которых очень большое. Применение теории планирования эксперимента может позволить получить набор моделей, основанных на реальных данных о вариациях конкретных мемристивных устройств и позволяющих оценить

точность работы ИНСМ с учетом взаимосвязи между параметрами сигналов задания сопротивления и весами ИНСМ, что согласуется с гипотезой, выдвинутой на основании результатов обзора и анализа литературы.

Данные результаты получены автором лично и были частично опубликованы в соавторстве (вклад автора более 50 %) в [100–103].

2 Разработка моделей и алгоритмов, позволяющих установить взаимосвязь между параметрами сигналов задания сопротивления и функциональной корректностью искусственных нейронных сетей на базе мемристоров

2.1 Модель и алгоритм моделирования зависимости сопротивления мемристивного устройства от параметров сигналов его задания

Модель 1. Модель зависимости сопротивления мемристивного устройства от параметров сигналов его задания:

$$\tilde{R} = f(F_1, \ldots, F_e), \quad (2.1)$$

$$\sigma_R = k(F_1, \ldots, F_e), \quad (2.2)$$

$$\sigma_R = s(\tilde{R}), \quad (2.3)$$

$$R = N(\tilde{R}, \sigma_R^2), \quad (2.4)$$

где F – это параметр сигнала задания сопротивления R МУ (например, напряжение, длительность или количество импульсов);

$\tilde{R}$ – выборочное среднее сопротивления R МУ;

σ_R – среднее квадратическое отклонение (СКО) сопротивления R МУ;

e – индекс параметра сигнала задания сопротивления R МУ;

$f()$ – функция зависимости $\tilde{R}$ от F, получаемая в результате интерполяции;

$k()$ – функция зависимости σ_R от F, получаемая в результате интерполяции;

$s()$ – функция зависимости σ_R от $\tilde{R}$;

R – сопротивление МУ;

$N()$ – нормальный закон распределения.

Функции $f()$, $k()$ и $s()$ можно получить путем интерполирования соответствующих экспериментально полученных значений R. Функция $s()$ необходима для того, чтобы при моделировании весов синапсов W ИНСМ сохранить взаимосвязь между сопротивлением R, разбросом его значений σ_R и

параметрами сигнала задания сопротивления F. Сопротивление R в данном случае будет являться случайной величиной.

Алгоритм 1. Алгоритм моделирования зависимости сопротивления мемристивного устройства от параметров сигналов его задания (рисунок 2.1):

Рисунок 2.1 – Алгоритм 1

Шаг 1. Вводятся исходные данные: параметры сигнала записи сопротивления (факторы) F, влияющие на сопротивление мемристора R и уровни их значений.

Шаг 2. Выполняется формирование плана эксперимента (таблица 2.1).

Таблица 2.1 – Пример плана эксперимента

Номер эксперимента	**Амплитуда импульса, В**	**Количество импульсов**	**Длительность импульса, мс**	**Результат эксперимента**
1	u_1	n_1	t_1	R_1
2	u_2	n_2	t_1	R_2
…	…	…	…	…
N	u_N	n_N	t_N	R_N

Шаг 3. Проверяется условие $i \leq N$, где N – количество экспериментов. Если данное условие выполняется, то i увеличивается на 1, и выполняется переход к шагу 4, иначе выполняется переход к шагу 8.

Шаг 4. Проверяется условие $j \leq M$, где M – количество параллельных опытов. Если данное условие выполняется, то j увеличивается на 1, и выполняется переход к шагу 5, иначе выполняется переход к шагу 7.

Шаг 5. Выполняется подача на мемристор сигнала с параметрами F, соответствующими данному опыту.

Шаг 6. Накапливаются получаемые значения сопротивлений R мемристора, рассчитываемые из значений АЦП по формуле (3.4). Переход к шагу 4.

Шаг 7. Выполняется расчет выборочного среднего $\tilde{R}$ и СКО σ_R для полученных значений сопротивлений. Переход к шагу 3.

Шаг 8. После завершения цикла выполняется интерполяция зависимости выборочного среднего $\tilde{R}$ и СКО σ_R сопротивления мемристора от параметров сигнала F, изменяющего его сопротивление.

Шаг 9. Выполняется вывод значений параметров модели зависимости сопротивления R МУ от параметров сигналов его задания F.

Дополнительно можно делать интерполяцию функции, обратной функции $s()$. Таким образом, полученные с применением данного алгоритма модели могут быть использованы как для определения того, какое сопротивление R будет у МУ в результате подачи сигнала с выбранными значениями параметров F, так и для

определения требуемых значений данных параметров F сигнала программирования для достижения конкретного значения сопротивления R.

2.2 Модель и алгоритм моделирования зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса

Модель 2. Модель зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса:

$$\widetilde{w} = g(R), \quad (2.5)$$

$$\sigma_w = u(\widetilde{w}), \quad (2.6)$$

$$w = N(\widetilde{w}, \sigma_w^2), \quad (2.7)$$

где $g()$ – функция зависимости веса синапса нейрона W от сопротивления R, берущаяся из схемотехники;

$\widetilde{w}$ – выборочное среднее веса синапса нейрона;

σ_w – СКО веса синапса нейрона;

$u()$ – функция зависимости СКО веса синапса нейрона σ_w от его выборочного среднего $\widetilde{w}$, получаемая в результате интерполяции;

$N()$ – нормальный закон распределения;

W – вес синапса нейрона (случайная величина).

Функцию $u()$ можно получить путем интерполирования экспериментально полученных значений. Она необходима для того, чтобы при оценке ФК ИНСМ сохранить взаимосвязь между весом синапса W, разбросом его значений σ_w и сопротивлением R МУ. Вес W в данном случае будет являться случайной величиной.

Алгоритм 2. Алгоритм моделирования зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса (рисунок 2.2):

Шаг 1. Вводятся исходные данные: уровни значений сопротивлений мемристора R, для которых будет выполняться моделирование веса.

Шаг 2. Выполняется формирование плана эксперимента.

Шаг 3. Проверяется условие $i \leq N$, где N – количество экспериментов. Если данное условие выполняется, то i увеличивается на 1, и выполняется переход к шагу 4, иначе выполняется переход к шагу 8.

Шаг 4. Проверяется условие $j \leq M$, где M – количество параллельных опытов. Если данное условие выполняется, то j увеличивается на 1, и выполняется переход к шагу 5, иначе выполняется переход к шагу 7.

Шаг 5. Проводится моделирование весового коэффициента W в соответствии с формулой (2.5), описывающей функционирование выбранной типовой схемы, и моделью мемристора по формуле (2.4), с параметром R, соответствующим данному опыту.

Шаг 6. Накапливаются получаемые значения весов W. Переход к шагу 4.

Шаг 7. Выполняется расчет выборочного среднего $\widetilde{w}$ и СКО σ_w для полученных значений весов. Переход к шагу 3.

Шаг 8. После завершения цикла выполняется интерполяция зависимости выборочного среднего $\widetilde{w}$ и СКО σ_w веса синапса нейрона от сопротивления мемристивного устройства R.

Шаг 9. Выполняется вывод значений параметров модели зависимости веса синапса нейрона от сопротивления мемристивного устройства.

Дополнительно можно делать интерполяцию функции, обратной функции $g()$. Таким образом, полученные с применением данного метода модели могут быть использованы как для определения того, какой вес W будет у синапса нейрона при заданном сопротивлении R, так и для определения требуемого значения сопротивлении R для записи конкретного значения веса синапса нейрона W.

Рисунок 2.2 – Алгоритм 2

2.3 Алгоритм оценки функциональной корректности искусственных нейронных сетей на базе мемристоров

Алгоритм 3. Алгоритм оценки ФК ИНСМ (рисунок 2.3):

Шаг 1. Вводятся исходные данные: веса W и параметры архитектуры ИНС.

Шаг 2. Проверяется условие $i \leq N$, где N – количество экспериментов. Если данное условие выполняется, то i увеличивается на 1, и выполняется переход к шагу 3, иначе выполняется переход к шагу 12.

Шаг 3. Проверяется условие $j \leq L$, где L – количество слоев ИНС. Если данное условие выполняется, то j увеличивается на 1, и выполняется переход к шагу 4, иначе выполняется переход к шагу 11.

Шаг 4. Проверяется условие: если данный слой имеет синапсы, реализованные на базе мемристоров, то переходим к шагу 5, иначе переходим к шагу 3.

Шаг 5. Выполняется умножение входных данных X на коэффициент K, для масштабирования.

Шаг 6. Выполняется моделирование веса W синапса нейрона по формуле (2.7).

Шаг 7. Выполняется матричное умножение входных данных X на значения весов синапсов W.

Шаг 8. Выполняется деление результата матричного умножения на коэффициент K.

Шаг 9. Выполняется функция активации нейронов слоя.

Шаг 10. К выходу нейронов применяется пороговая функция (2.8) со значением порога T, Выполняется Переход к шагу 3.

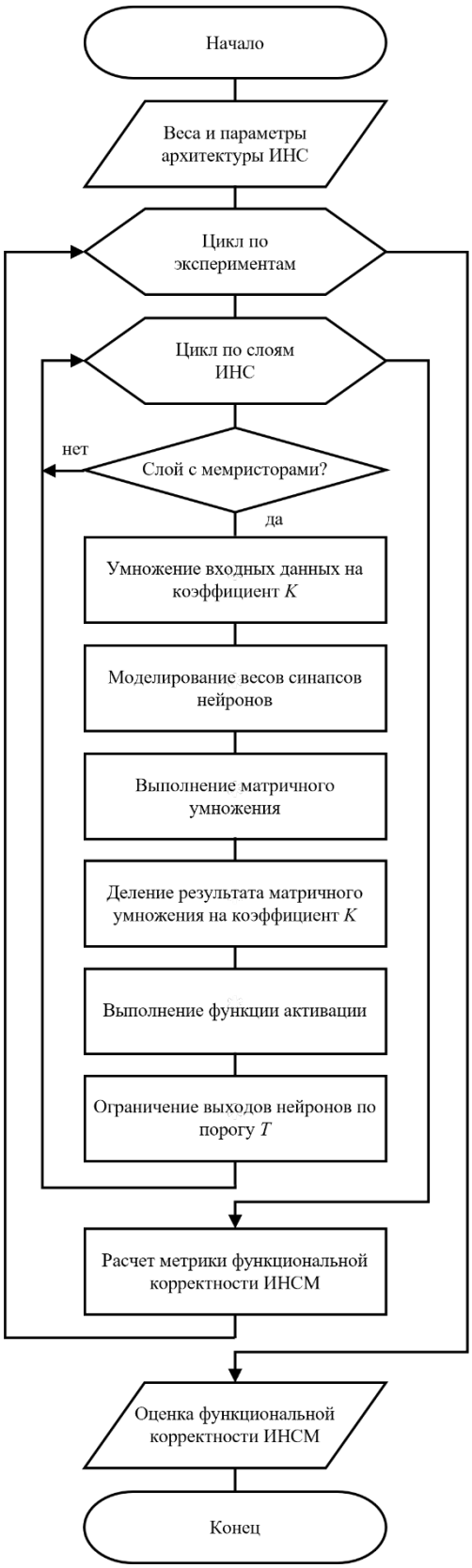


Рисунок 2.3 – Алгоритм 3

Шаг 11. Выполняется расчет метрики оценки L ФК ИНСМ для данного эксперимента. Переход к шагу 2.

Шаг 12. Выполняется вывод значений метрики оценки L ФК ИНСМ.

$$y_{i,j} = \begin{cases} y_{i,j}, y_{i,j} \leq T \\ T, y_{i,j} > T \end{cases}, \quad (2.8)$$

где $y_{i,j}$ – выход i слоя j нейрона;

T – максимально допустимое рабочее напряжение, подаваемое на вход кроссбар-массива мемристивных устройств.

Расчет масштабирующего коэффициента K выполняется аналитически или экспериментально. Для аналитического расчета K необходимо выполнить следующее:

1) Масштабировать тестовую выборку X до диапазона максимального рабочего напряжения мемристоров T по формуле

$$V_{in} = \frac{x}{max(x)} T, \quad (2.9)$$

где X – данные тестовой выборки;

max(X) – максимальное значение в тестовой выборке.

2) Выполнить компьютерное моделирование работы ИНСМ по алгоритму 3, сохраняя выходные данные с каждого слоя модели отдельно.

3) Для сохраненных выходных данных нужно найти максимальное значение $max(y_{out}^{L-1})$ по слоям и рассчитать K_L для соответствующего слоя L по формуле (2.10).

$$K_L = \frac{max(y_{out}^{L-1})}{T}, \quad (2.10)$$

где K_L – коэффициент масштабирования выхода L-го слоя.

Если при введении в модель аналитически рассчитанного значения коэффициента масштабирования K ухудшилось значение метрики ФК ИНСМ при ее аппаратной реализации, то это значение можно подобрать экспериментально. Это связано с тем, что для слишком низких значений сопротивлений соотношение сигнал/шум уменьшается и шумы, присутствующие в обрабатываемых сигналах, начинают оказывать большее влияние на погрешность матричного умножения. Для определения коэффициента масштабирования K с применением методов теории планирования эксперимента необходимо выполнить следующее:

1) Задать диапазон значений и уровни масштабирующего коэффициента K.

2) Сформировать план эксперимента.

3) Выполнить компьютерное моделирование работы ИНСМ для разных значений масштабирующего коэффициента K с применением пороговой функции (2.8) и оценить ФК ИНСМ в соответствии с выбранной метрикой L.

4) Выбрать то значение масштабирующего коэффициента K, для которого используемая метрика ФК ИНСМ имеет наилучшее значение.

Влияние коэффициента K на точность ИНСМ проиллюстрировано на рисунке 2.4.

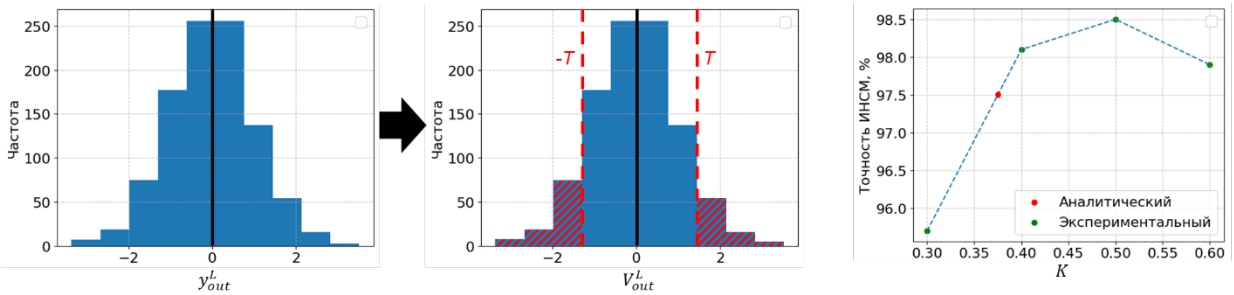


Рисунок 2.4 – Иллюстрация влияния коэффициента K на точность ИНСМ: а) иллюстрация значимости влияния T; б) иллюстрация сравнения аналитического и экспериментального подходов к поиску K

2.4 Исследование предложенных моделей и алгоритмов моделирования

2.4.1 Описание модельной среды

Для предварительного исследования разработанных моделей и алгоритмов моделирования было принято решение использовать не реальные мемристивные устройства и вычислители, а создать модельную среду, в которой разработанные алгоритмы будут реализованы программно, а аппаратная часть будет подменена схемотехнической моделью. Это позволит с одной стороны подготовить все необходимые программные средства для дальнейшей апробации на реальных устройствах, а с другой стороны провести исследование и детальную демонстрацию работы предложенных алгоритмов.

Процесс создания модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания обязательно включает этап задания сопротивления. Задание различных значений сопротивлений является одной их основных функциональных возможностей мемристивных устройств, играющей важную роль в точности вычислений, поскольку оказывает влияние на значения весовых коэффициентов синапсов нейронов.

Все возможные параметры сигнала, влияющие на сопротивление описаны в п. 1.5. Однако стоит отметить, что для обеспечения возможности изменения всех этих параметров в процессе записи сопротивления приведет к значительному усложнению схемы устройства. В идеальном случае должна быть обеспечена возможность записи нужного сопротивления импульсами, в которых меняется только один параметр. В статье [104] для этого изменяли один параметр N, но при таком подходе запись весов отличается по времени. В данной главе рассмотрим более быстрый вариант записи, подавая на мемристивное устройство импульсы с одинаковой длительностью, но разной амплитудой при N=1.

Для построения модели мемристивного устройства и веса, по алгоритмам, описанным в пункте 2.1 и 2.2, было разработано специальное программное обеспечение на языке программирования Python. Данный язык программирования выбран исходя из его кроссплатформенности и большого количества библиотек для выполнения различных задач, таких как интерполяция данных, их визуализация и т.д., что позволяет существенно упростить процесс разработки и расширить пользовательскую базу. Данная программа реализует алгоритм 1 и 2 и имеет следующие возможности:

– Визуализация экспериментальных данных используемых для создания моделей, графиков интерполяции.

– Оценка МО и СКО экспериментальных данных.

– Интерполяция экспериментальных данных для мемристивного устройства и весового коэффициента, на основе схемы из одного или двух мемристоров.

– Вывод формул, полученных в результате интерполяции.

– Вывод таблиц с данными по интерполяции.

Кроме того, в данном программном обеспечении реализован функционал взаимодействия с системой для схемотехнического моделирования LTSpice, что позволяет получать экспериментальные данные о задании сопротивления мемристивного устройства на основе данных Spice моделирования. В данной главе все эксперименты будут выполняться на модели, представленной на рисунке 2.5.

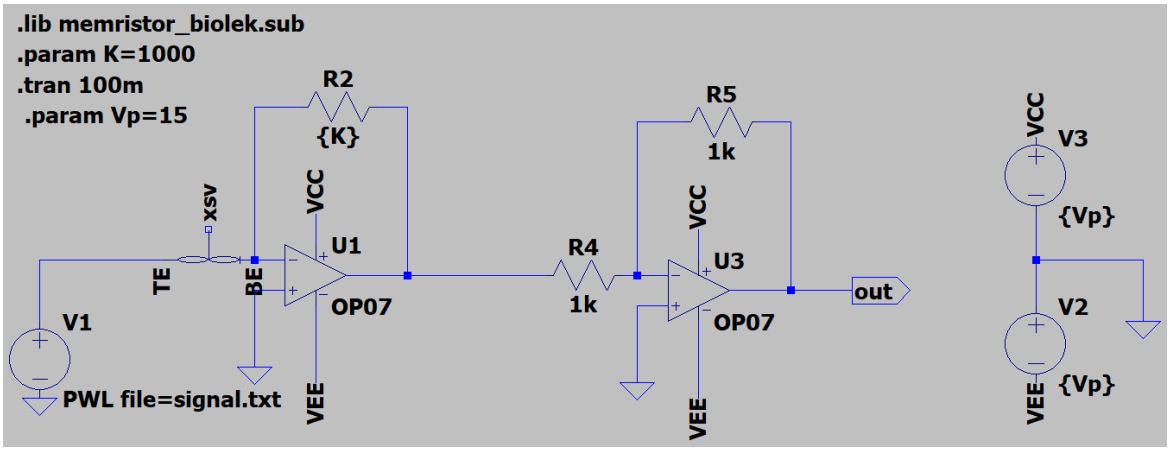


Рисунок 2.5 – Схема задания сопротивления мемристивного устройства

Как видно из представленной выше схемы в качестве генератора сигнала использовался источник напряжения V1, который генерирует необходимый сигнал в соответствии с файлом signal.txt, в котором заданы пары значений (время, напряжение). Для преобразования тока в цепи с мемристором в напряжение использовались два операционных усилителя U1 и U3, подключенных последовательно. Коэффициент усиления трансимпедансного усилителя U1 задается константой K, коэффициент усиления инвертирующего усилителя U3 равен -1, он используется для смены знака напряжения на выходе. В качестве мемристивного устройства использовалась модель «Biolek» [105]. При этом вариации в модель добавлялась путем изменения значения максимального и минимального сопротивления мемристивного устройства по нормальному закону распределения.

Для выполнения алгоритма моделирования зависимости сопротивления мемристивного устройства от параметров сигналов его задания требуется многократное изменение сопротивления мемристивного устройства с разными параметрами сигнала и накопление статистики, что влечет за собой существенные временные затраты в связи с отсутствием возможности работы LTSpice в

многопроцессном режиме. Для ускорения данной процедуры в программу была внедрена специальная функция, которая создает несколько копий данной программы со всеми необходимыми параметрами, включая схемы моделирования, и запускает их параллельно в разных процессах, что позволяет существенно сократить время получения экспериментальных данных от модели.

2.4.2 Создание модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания

Создание модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания проведено для однофакторного эксперимента, в котором время импульса равно 100 миллисекундам, количество данных импульсов не может быть больше единицы в одном цикле, при этом есть возможность изменять амплитуду импульса в диапазоне от 0,1 В до 3,1 В включительно. Разброс минимального и максимального сопротивления мемристивного устройства установлен на уровне 5 %. В соответствии с этим был сформирован план эксперимента, представленный в таблице 2.2.

Таблица 2.2 – План эксперимента для алгоритма 1

Номер эксперимента	**Амплитуда импульса, В**	**Количество импульсов**	**Длительность импульса, мс**
1	0,1	1	100
2	0,43	1	100
3	0,77	1	100
4	1,1	1	100
5	1,43	1	100
6	1,77	1	100
7	2,1	1	100
8	2,43	1	100
9	2,77	1	100
10	3,1	1	100

Затем выполнялся цикл моделирования, где количество параллельных экспериментов при выполнении алгоритма 1 было установлено в 1000. Подача на мемристор сигнала осуществлялось следующим образом:

– Перевод мемристивного устройства в высокомное состояние двумя последовательными импульсами по 100 миллисекунд с амплитудой -3,1 В.

– Подача сигнала на мемристивное устройство в соответствии с планом эксперимента.

В результате выполнения алгоритма 1 были визуализированы накопленные значения сопротивлений *R*, представленные на рисунке 2.6.

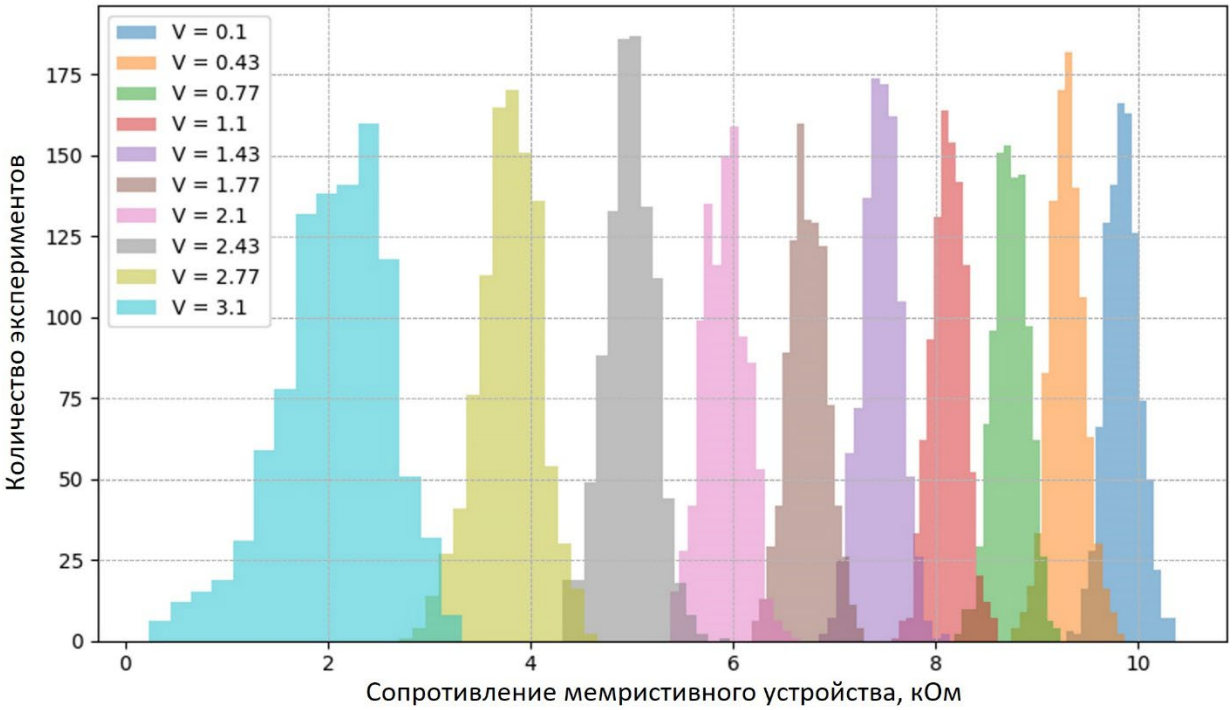


Рисунок 2.6 – Гистограммы разброса сопротивлений мемристивного устройства

В таблице 2.3 представлена оценка МО и СКО сопротивлений *R*.

Таблица 2.3 – Статистика полученных данных

Напряжение, В	**МО сопротивления мемристора, кОм**	**СКО сопротивления мемристора, кОм**
0,1	9,85	0,17
0,43	9,3	0,17
0,77	8,74	0,17
1,1	8,12	0,17
1,43	7,46	0,19
1,77	6,73	0,19
2,1	5,92	0,21
2,43	4,98	0,24
2,77	3,79	0,3
3,1	2,05	0,55

Из полученных данных можно сделать вывод, что использование данного диапазона напряжений позволяет покрыть большую часть диапазона

сопротивлений мемристивного устройства. Важно также отметить, что наименьшая погрешность достигается при меньших входных напряжениях, и соответственно, чем меньше сопротивление мемристивного устройства, тем больше погрешность и нелинейность процесса задания сопротивления.

Для полученных данных МО и СКО, которые представлены в таблице 2.3, в соответствии с алгоритмом 1 была выполнена интерполяция с применением библиотеки SciPy и функции interp1d. В данном случае было использовано два вида интерполяции, а именно кусочно-линейная (КЛИ) и кусочно-кубическая интерполяция (ККИ).

Пример модели кусочно-линейной интерполяции приведен в виде формул (2.11, 2.12).

$$\tilde{R} = \begin{cases} -1{,}65V + 10{,}01 \text{ при } 0{,}1 \leqslant V \leqslant 0{,}43 \\ -1{,}67V + 10{,}02 \text{ при } 0{,}43 \leqslant V \leqslant 0{,}77 \\ -1{,}84V + 10{,}15 \text{ при } 0{,}77 \leqslant V \leqslant 1{,}1 \\ -1{,}98V + 10{,}31 \text{ при } 1{,}1 \leqslant V \leqslant 1{,}43 \\ -2{,}19V + 10{,}6 \text{ при } 1{,}43 \leqslant V \leqslant 1{,}77 \\ -2{,}44V + 11{,}04 \text{ при } 1{,}77 \leqslant V \leqslant 2{,}1 \\ -2{,}82V + 11{,}84 \text{ при } 2{,}1 \leqslant V \leqslant 2{,}43 \\ -3{,}58V + 13{,}7 \text{ при } 2{,}43 \leqslant V \leqslant 2{,}77 \\ -5{,}21V + 18{,}2 \text{ при } 2{,}77 \leqslant V \leqslant 3{,}1 \end{cases}, \quad (2.11)$$

где V– амплитуда сигнала.

$$\sigma_R = \begin{cases} 0{,}01V + 0{,}17 \text{ при } 0{,}1 \leqslant V \leqslant 0{,}43 \\ -0{,}01V + 0{,}17 \text{ при } 0{,}43 \leqslant V \leqslant 0{,}77 \\ -0{,}01V + 0{,}18 \text{ при } 0{,}77 \leqslant V \leqslant 1{,}1 \\ 0{,}07V + 0{,}09 \text{ при } 1{,}1 \leqslant V \leqslant 1{,}43 \\ 0{,}01V + 0{,}17 \text{ при } 1{,}43 \leqslant V \leqslant 1{,}77 \\ 0{,}06V + 0{,}08 \text{ при } 1{,}77 \leqslant V \leqslant 2{,}1 \\ 0{,}08V + 0{,}05 \text{ при } 2{,}1 \leqslant V \leqslant 2{,}43 \\ 0{,}18V - 0{,}2 \text{ при } 2{,}43 \leqslant V \leqslant 2{,}77 \\ 0{,}74V - 1{,}75 \text{ при } 2{,}77 \leqslant V \leqslant 3{,}1 \end{cases}, \quad (2.12)$$

Далее была выполнена оценка относительной погрешности модельных данных. Данная оценка осуществлялась в точках, расположенных ровно посередине между теми данными, которые использовались для вычисления коэффициентов функции интерполяции (обучающие данные). Для этих проверочных точек с мемристивного устройства также были получены соответствующие данные, такие как МО и СКО.

Результаты работы алгоритма и проверки моделей представлены на рисунке 2.7 и в таблице 2.4 и 2.5.

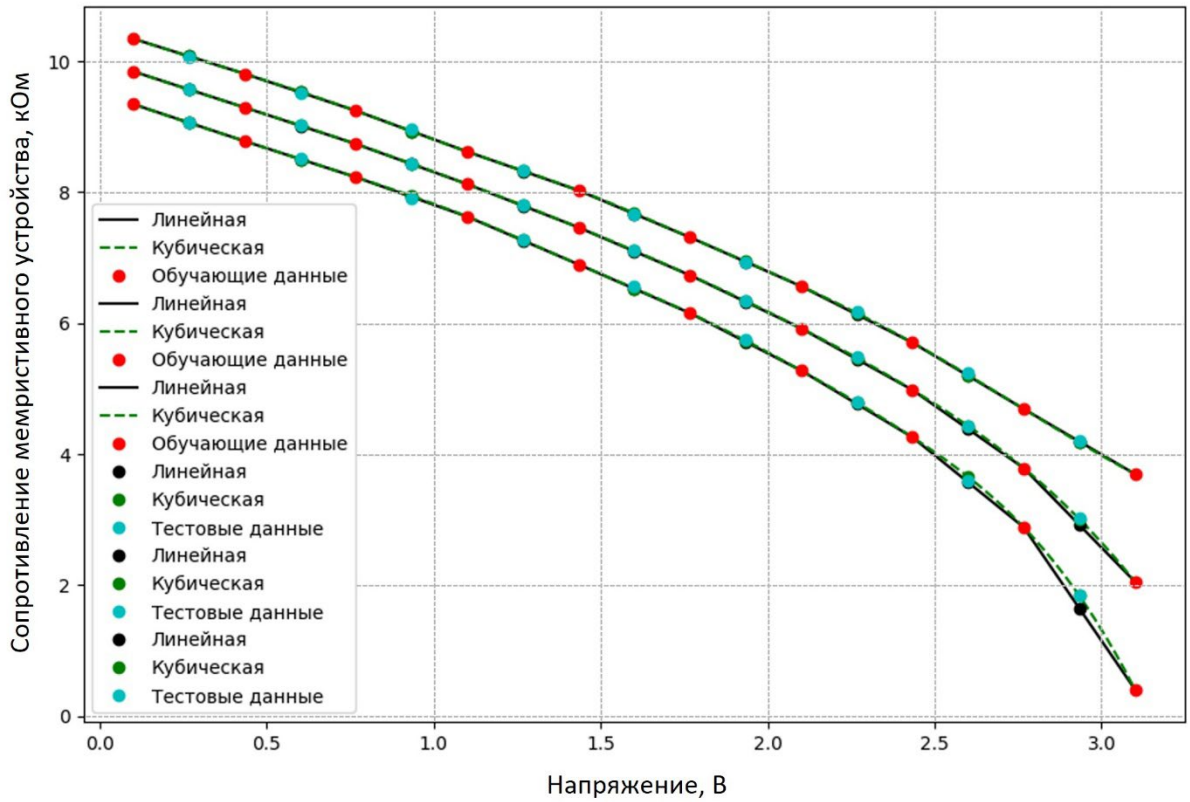


Рисунок 2.7 – Тестирование моделей интерполяции

Таблица 2.4 – КЛИ модель мемристивного устройства по данным МО и СКО

V	МО, кОм	СКО, %	КЛИ МО, кОм	КЛИ СКО, %
0,27	9,58	5,15	9,57 (0,06%)	5,3 (2,91%)
0,6	9,03	5,66	9,02 (0,09%)	5,66 (0,0%)
0,93	8,43	6,07	8,43 (0,01%)	5,95 (1,98%)
1,27	7,81	6,74	7,79 (0,14%)	6,8 (0,89%)
1,6	7,11	7,85	7,1 (0,1%)	8,05 (2,55%)
1,93	6,33	9,3	6,33 (0,04%)	9,66 (3,87%)
2,27	5,47	12,28	5,45 (0,33%)	12,5 (1,79%)
2,6	4,43	18,51	4,38 (1,16%)	18,34 (0,92%)
2,93	3,05	37,2	2,92 (4,34%)	41,81 (1,23%)

Из полученных данных видно, что в большинстве тестовых точек полученные модели имеют относительную погрешность интерполяции (представлена в скобках в таблице) МО меньше 0,5%, однако, при приближении значения к напряжению 3 В можно наблюдать тенденцию увеличения данной погрешности для обоих типов интерполяций. Это связано с тем, что при более низких значениях сопротивления, мемристивное устройство ведет себя сильно

нелинейно и для его более точного моделирования в данной области требуется большее число точек. В итоге максимальная относительная ошибка линейной интерполяции МО составляет 4,34%, а кубической 1,44%.

Таблица 2.5 – ККИ модель мемристивного устройства по данным МО и СКО

V	МО, кОм	СКО, %	ККИ МО, кОм	ККИ СКО, %
0,27	9,58	5,15	9,57 (0,09%)	5,31 (3,11%)
0,6	9,03	5,66	9,02 (0,03%)	5,71 (0,88%)
0,93	8,43	6,07	8,44 (0,09%)	5,86 (3,46%)
1,27	7,81	6,74	7,8 (0,06%)	6,79 (0,74%)
1,6	7,11	7,85	7,11 (0,03%)	8,08 (2,93%)
1,93	6,33	9,3	6,34 (0,15%)	9,55 (2,69%)
2,27	5,47	12,28	5,47 (0,05%)	12,49 (1,71%)
2,6	4,43	18,51	4,43 (0,09%)	17,53 (3,29%)
2,93	3,05	37,2	3,01 (1,44%)	38,39 (3,2%)

Если же рассматривать СКО, то максимальная ошибка линейной интерполяции СКО составляет 3,87%, а кубической 3,29%.

По итогу можно сказать, что кусочно-кубическая интерполяция немного лучше справляется с задачей интерполяции данных, однако при этом имеет более высокую вычислительную сложность.

Далее выполнена кусочно-линейная и кусочно-кубическая интерполяция V от $\tilde{R}$. Пример формулы кусочно-линейной интерполяционной модели, представлен ниже:

$$V = \begin{cases} -0{,}19\tilde{R} + 3{,}49 \text{ при } 2{,}05 \leqslant \tilde{R} \leqslant 3{,}79 \\ -0{,}28\tilde{R} + 3{,}82 \text{ при } 3{,}79 \leqslant \tilde{R} \leqslant 4{,}98 \\ -0{,}35\tilde{R} + 4{,}2 \text{ при } 4{,}98 \leqslant \tilde{R} \leqslant 5{,}92 \\ -0{,}41\tilde{R} + 4{,}53 \text{ при } 5{,}92 \leqslant \tilde{R} \leqslant 6{,}73 \\ -0{,}46\tilde{R} + 4{,}84 \text{ при } 6{,}73 \leqslant \tilde{R} \leqslant 7{,}46, \\ -0{,}5\tilde{R} + 5{,}2 \text{ при } 7{,}46 \leqslant \tilde{R} \leqslant 8{,}12 \\ -0{,}54\tilde{R} + 5{,}51 \text{ при } 8{,}12 \leqslant \tilde{R} \leqslant 8{,}74 \\ -0{,}6\tilde{R} + 5{,}99 \text{ при } 8{,}74 \leqslant \tilde{R} \leqslant 9{,}3 \\ -0{,}61\tilde{R} + 6{,}07 \text{ при } 9{,}3 \leqslant \tilde{R} \leqslant 9{,}85 \end{cases} \quad (2.13)$$

Далее была выполнена оценка относительной погрешности модельных данных. Данная оценка, также как и в предыдущем случае, была осуществлена в точках, расположенных ровно посередине между теми данными, которые использовались для вычисления коэффициентов функции интерполяции.

Результаты работы алгоритма и проверки моделей представлены на рисунке 2.8 и в таблице 2.6.

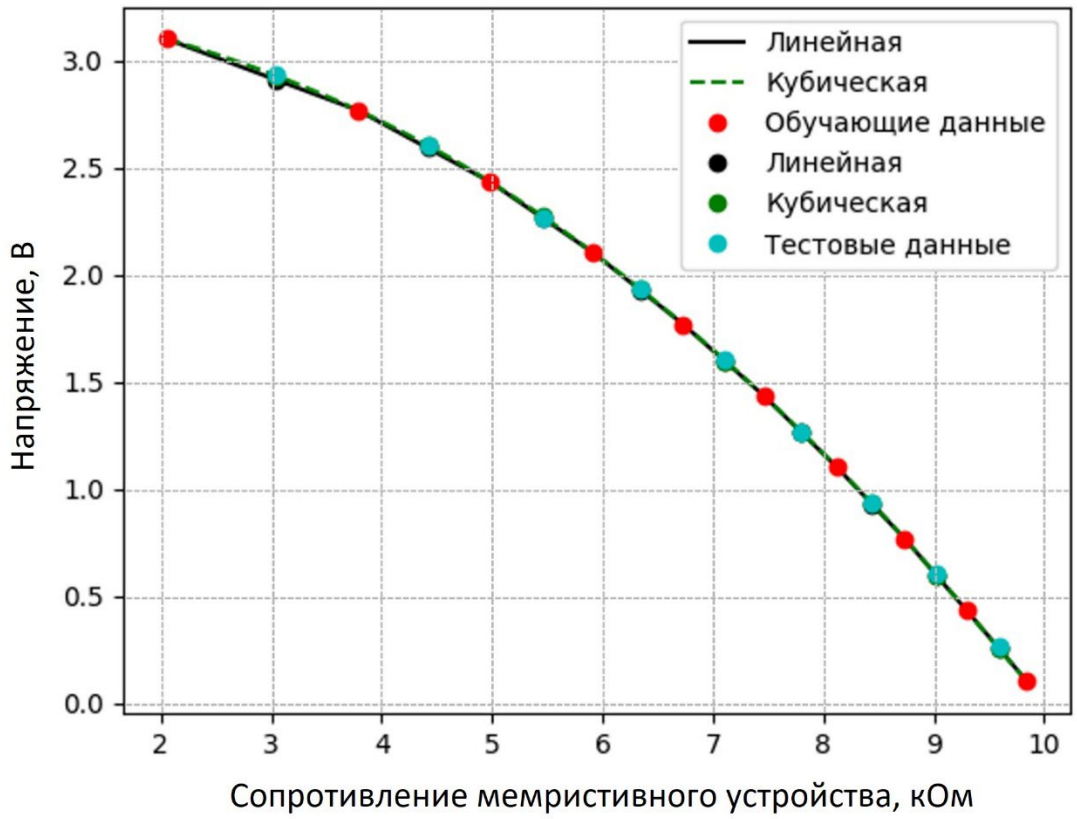


Рисунок 2.8 – Тестирование моделей мемристивного устройства

Таблица 2.6 – Сравнение точности моделей КЛИ и ККИ

$\widetilde{R}$, кОм	Экспериментальные данные, В	КЛИ, В	ККИ, В
3,04	2,93	2,91 (0,8%)	2,93 (0,14%)
4,42	2,6	2,59 (0,37%)	2,6 (0,09%)
5,46	2,27	2,26 (0,17%)	2,27 (0,16%)
6,34	1,93	1,93 (0,29%)	1,93 (0,03%)
7,11	1,6	1,59 (0,35%)	1,6 (0,07%)
7,8	1,27	1,26 (0,43%)	1,26 (0,18%)
8,44	0,93	0,93 (0,64%)	0,93 (0,22%)
9,02	0,6	0,6 (0,64%)	0,6 (0,19%)
9,59	0,27	0,26 (4,26%)	0,25 (4,81%)

Из полученных данных видно, что в большинстве тестовых точек полученные модели имеют погрешность интерполяции меньше 0,5%, однако, при приближении значения к напряжению 9,59 кОм можно наблюдать тенденцию увеличения данной погрешности для обоих типов интерполяций. В итоге максимальная ошибка линейной интерполяции составляет 4,26%, а кубической

4,81%. По итогу можно сказать, что кусочно-кубическая интерполяция немного хуже справляется с задачей интерполяции данных.

2.4.3 Создание модели зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса

Допустим, что синапс ИНСМ аппаратно реализован, в виде схемы из двух мемристоров, представленной на рисунке 2.9. Формула, по которая формирует вес в соответствии с представленной схемой представлена ниже:

$$W = R_f \frac{R_{M1} - R_{M2}}{R_{M1} R_{M2}}, \tag{2.14}$$

где W – вес синапса;

R_{M1}, R_{M2} – сопротивления мемристоров;

R_f – сопротивление обратной связи.

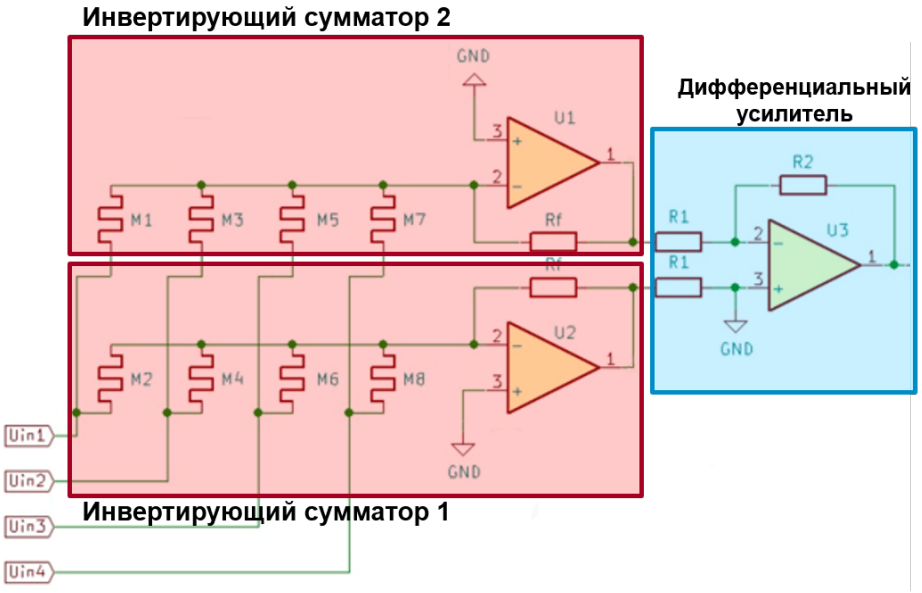


Рисунок 2.9 – Схема веса синапса нейрона

Сопротивление обратной связи R_f в данном случае равно 10 кОм. В качестве номинального значения R_{M2} используется максимальное интерполируемое сопротивление, то есть 9,59 кОм. Значение R_{M1} берется из диапазона от 2,1 до 9,01, поделённого на 9 равных участков.

На основе представленных входных данных сформирован план эксперимента, представленный в таблице 2.7.

Таблица 2.7 – План эксперимента для алгоритма 2

Номер эксперимента	R_{M1}, кОм	R_{M2}, кОм	R_f, кОм
1	2,1	9,59	10
2	2,96	9,59	10
3	3,83	9,59	10
4	4,69	9,59	10
5	5,55	9,59	10
6	6,42	9,59	10
7	7,28	9,59	10
8	8,14	9,59	10
9	9,01	9,59	10

Затем выполнялся цикл моделирования, в котором количество параллельных экспериментов при выполнении алгоритма 2 было установлено в 1000. В результате выполнения алгоритма 2 визуализированы накопленные значения сопротивлений *W*, представленные на рисунке 2.10.

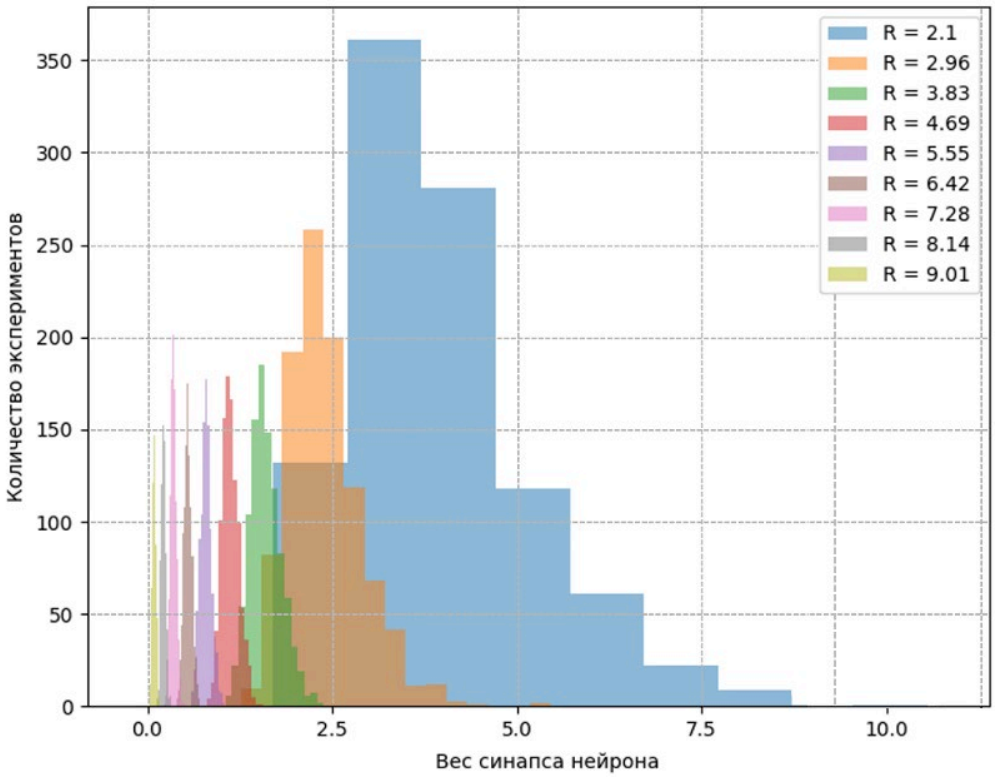


Рисунок 2.10 – Гистограммы формирования веса синапса нейрона из мемристивных устройств

В таблице 2.8 представлена оценка МО и СКО веса *W*.

Таблица 2.8 – Статистика данных формирования веса

Сопротивление мемристора, кОм	**МО веса**	**СКО веса**
2,1	4,06	1,61
2,96	2,41	0,5
3,83	1,6	0,21
4,69	1,12	0,12
5,55	0,79	0,07
6,42	0,55	0,05
7,28	0,36	0,04
8,14	0,22	0,03
9,01	0,1	0,03

Для полученных данных МО и СКО, которые представлены в таблице 2.8, в соответствии с алгоритмом 2 была выполнена интерполяция с применением библиотека SciPy и функции interp1d. В данном случае было использовано два вида интерполяции, а именно кусочно-линейная и кусочно-кубическая интерполяция.

Пример модели кусочно-линейной интерполяции приведен в виде формул (2.15, 2.16)

$$\widetilde{w} = \begin{cases} -1{,}91R + 8{,}07 \text{ при } 2{,}1 \leqslant R \leqslant 2{,}96 \\ -0{,}94R + 5{,}21 \text{ при } 2{,}96 \leqslant R \leqslant 3{,}83 \\ -0{,}55R + 3{,}71 \text{ при } 3{,}83 \leqslant R \leqslant 4{,}69 \\ -0{,}38R + 2{,}93 \text{ при } 4{,}69 \leqslant R \leqslant 5{,}55 \\ -0{,}28R + 2{,}36 \text{ при } 5{,}55 \leqslant R \leqslant 6{,}42 \\ -0{,}22R + 1{,}94 \text{ при } 6{,}42 \leqslant R \leqslant 7{,}28 \\ -0{,}17R + 1{,}6 \text{ при } 7{,}28 \leqslant R \leqslant 8{,}14 \\ -0{,}14R + 1{,}34 \text{ при } 8{,}14 \leqslant R \leqslant 9{,}01 \end{cases}. \quad (2.15)$$

$$\sigma_w = \begin{cases} -1{,}28R + 4{,}31 \text{ при } 2{,}1 \leqslant R \leqslant 2{,}96 \\ -0{,}33R + 1{,}49 \text{ при } 2{,}96 \leqslant R \leqslant 3{,}83 \\ -0{,}11R + 0{,}65 \text{ при } 3{,}83 \leqslant R \leqslant 4{,}69 \\ -0{,}05R + 0{,}35 \text{ при } 4{,}69 \leqslant R \leqslant 5{,}55 \\ -0{,}02R + 0{,}21 \text{ при } 5{,}55 \leqslant R \leqslant 6{,}42 \\ -0{,}01R + 0{,}14 \text{ при } 6{,}42 \leqslant R \leqslant 7{,}28 \\ -0{,}01R + 0{,}11 \text{ при } 7{,}28 \leqslant R \leqslant 8{,}14 \\ -0{,}01R + 0{,}08 \text{ при } 8{,}14 \leqslant R \leqslant 9{,}01 \end{cases}. \quad (2.16)$$

Далее была выполнена оценка относительной погрешности модельных данных. Данная оценка выполнялась в точках, расположенных ровно посередине между теми данными, которые использовались для вычисления коэффициентов функции интерполяции. Для этих проверочных точек с модели весового коэффициента, в качестве которого выступает формула расчета веса, также были получены соответствующие данные, такие как МО и СКО.

Результаты работы и проверки моделей представлены на рисунке 2.11 и в таблице 2.9 и 2.10.

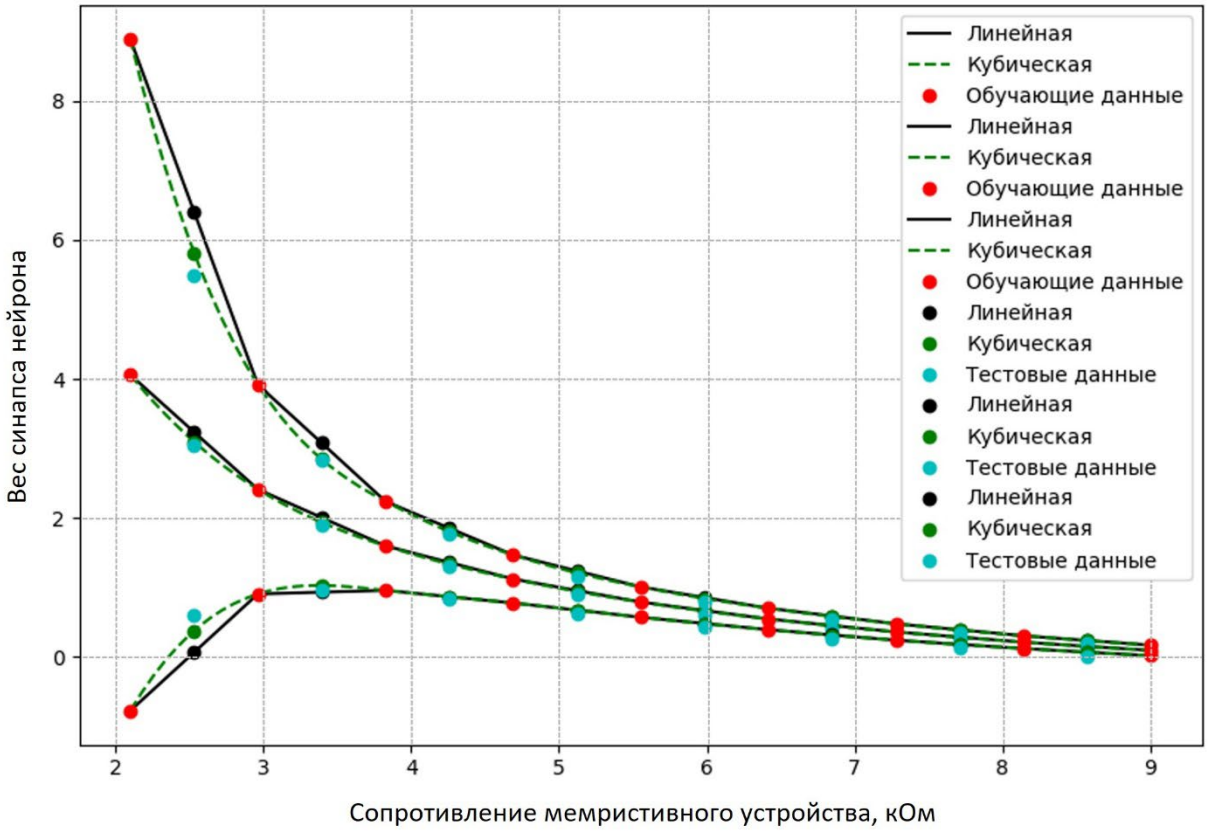


Рисунок 2.11 – Тестирование моделей весового коэффициента

Таблица 2.9 – Точность модели КЛИ весового коэффициента по данным МО и СКО

Сопротивление, кОм	**МО веса**	**СКО веса, %**	**КЛИ МО**	**КЛИ СКО, %**
2,53	3,05	79,92	3,24 (3,21%)	103,96 (20,08%)
3,4	1,9	48,69	2,01 (3,56%)	56,43 (15,9%)
4,26	1,3	36,42	1,36 (3,47%)	37,8 (3,79%)
5,12	0,89	30,25	0,96 (7,25%)	31,55 (4,3%)
5,98	0,61	29,91	0,67 (4,18%)	30,62 (2,37%)
6,85	0,4	32,9	0,46 (2,99%)	34,09 (3,62%)
7,71	0,24	43,2	0,29 (4,97%)	43,92 (1,67%)
8,57	0,11	82,17	0,16 (4,31%)	78,85 (4,04%)

Из полученных данных видно, что в большинстве тестовых точек полученные модели имеют погрешность интерполяции МО больше 4 %, однако, при приближении значения к сопротивлению 8,57 кОм можно наблюдать тенденцию увеличения данной погрешности для обоих типов интерполяций. В итоге максимальная ошибка линейной интерполяции МО составляет 7,25%, а

кубической 5,73%. Относительная ошибка линейной интерполяции СКО составляет 20,08%, а кубической 5,25%, что лучше, чем для МО.

По итогу можно сказать, что кубическая интерполяция лучше справляется с задачей интерполяции данных по весовому коэффициенту синапса.

Анализируя полученные в таблице 2.9 данные можно сказать, что всего лишь 5% погрешность мемристивного устройства приводит практически к 30 % и более погрешности весового коэффициента, что еще раз подчеркивает важность проведения оценки ФК ИНСМ на моделях на этапе проектирования, чтобы с меньшей вероятность попасть в ситуацию, когда разработанное устройство не будет удовлетворять предъявляемых ему требований.

Таблица 2.10 – Точность модели ККИ весового коэффициента по данным МО и СКО

Сопротивление, кОм	**МО веса**	**СКО веса, %**	**ККИ МО**	**ККИ СКО, %**
2,53	3,05	79,92	3,1 (1,63%)	89,22 (5,64%)
3,4	1,9	48,69	1,94 (1,95%)	47,71 (2,01%)
4,26	1,3	36,42	1,34 (2,58%)	35,96 (1,26%)
5,12	0,89	30,25	0,94 (5,73%)	29,92 (1,09%)
5,98	0,61	29,91	0,66 (3,74%)	29,91 (0,0%)
6,85	0,4	32,9	0,45 (2,54%)	33,56 (2,01%)
7,71	0,24	43,2	0,28 (4,2%)	43,42 (0,51%)
8,57	0,11	82,17	0,15 (3,53%)	77,86 (5,25%)

Из представленных в таблице 2.10 данных можно заметить одну закономерность, которая заключается в том, что при приближении к границам диапазона сопротивлений погрешность веса увеличивается. В случае больших сопротивлений это связано с маленькими значениями весового коэффициента, а в случае маленьких сопротивлений связано с высокой нелинейность мемристивного диапазона на этом участке, что подтверждается данными графика 2.11.

Далее выполнена кусочно-линейная и кусочно-кубическая интерполяция R от $\tilde{R}$. Пример формулы кусочно-линейной интерполяционной модели, представлен ниже:

$$R = \begin{cases} -8{,}96\tilde{w} + 9{,}87 \text{ при } 0 \leqslant \tilde{w} \leqslant 0{,}1 \\ -7{,}25\tilde{w} + 9{,}7 \text{ при } 0{,}1 \leqslant \tilde{w} \leqslant 0{,}22 \\ -5{,}89\tilde{w} + 9{,}41 \text{ при } 0{,}22 \leqslant \tilde{w} \leqslant 0{,}36 \\ -4{,}6\tilde{w} + 8{,}94 \text{ при } 0{,}36 \leqslant \tilde{w} \leqslant 0{,}55 \\ -3{,}55\tilde{w} + 8{,}37 \text{ при } 0{,}55 \leqslant \tilde{w} \leqslant 0{,}79 \\ -2{,}6\tilde{w} + 7{,}61 \text{ при } 0{,}79 \leqslant \tilde{w} \leqslant 1{,}12 \\ -1{,}81\tilde{w} + 6{,}73 \text{ при } 1{,}12 \leqslant \tilde{w} \leqslant 1{,}6 \\ -1{,}06\tilde{w} + 5{,}52 \text{ при } 1{,}6 \leqslant \tilde{w} \leqslant 2{,}41 \end{cases} \quad (2.17)$$

Далее была выполнена оценка ошибки полученных моделей. Данная оценка, также как и в предыдущем случае, выполнялась в точках, расположенных ровно посередине между теми данными, которые использовались для вычисления коэффициентов функции интерполяции. Для этих проверочных точек с модели веса также были получены соответствующие данные, такие как МО и СКО.

Результаты работы и проверки моделей представлены на рисунке 2.12 и в таблице 2.11.

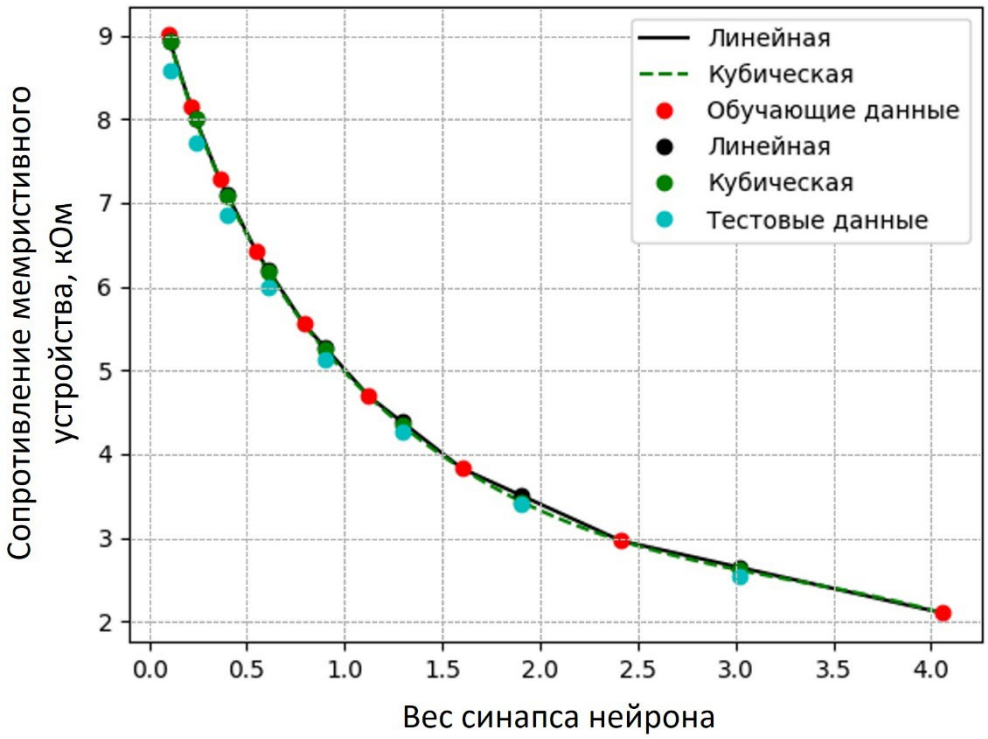


Рисунок 2.12 – Тестирование моделей весового коэффициента

Из полученных данных видно, что в большинстве тестовых точек полученные модели имеют погрешность интерполяции МО меньше 5%, однако, при приближении значений к максимальному и минимальному можно наблюдать тенденцию увеличения данной погрешности для обоих типов интерполяций, что

обусловлено причинами, описанными ранее в предыдущих моделях. В итоге максимальная ошибка линейной интерполяции МО составляет 4,55%, а кубической 4,07%.

Таблица 2.11 – Сравнение точности моделей весового коэффициента по данным МО

МО веса	**Сопротивление мемристора, кОм**	**КЛИ сопротивления мемристора, кОм**	**ККИ сопротивления мемристора, кОм**
0,11	8,57	8,93 (4,16%)	8,92 (4,07%)
0,24	7,71	8,01 (3,88%)	8,0 (3,72%)
0,4	6,85	7,11 (3,77%)	7,09 (3,49%)
0,61	5,98	6,2 (3,56%)	6,17 (3,18%)
0,9	5,12	5,28 (3,06%)	5,24 (2,41%)
1,29	4,26	4,38 (2,97%)	4,34 (2,03%)
1,91	3,4	3,5 (3,14%)	3,43 (0,98%)
3,02	2,53	2,65 (4,55%)	2,62 (3,29%)

По итогу можно сказать, что кусочно-кубическая интерполяция немного лучше справляется с данной задачей.

2.7 Выводы по главе

1) Разработана модель и алгоритм моделирования зависимости сопротивления мемристивного устройства от параметров сигналов его задания. Полученная модель позволяет не только выполнять анализ данных, то есть предположить при каких параметрах сигнала какое получиться сопротивление, но и решать обратную задачу, то есть выполнять синтез, который заключается в определении того, какие параметры сигнала нужно установить, чтобы получить нужное сопротивление, и какая у него будет погрешность.

2) Разработана модель и алгоритм моделирования зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса, которая может применяться для моделирования ИНСМ и основана на экспериментальных данных о вариациях реально заданных сопротивлений. Полученная модель позволяет не только выполнять анализ данных, то есть

предположить при каких сопротивления какое получиться значение веса, но и решать обратную задачу, то есть выполнять синтез, который заключается в определении того какие сопротивления нужно установить, чтобы получить нужное значение веса, и какая у него будет погрешность. Причем объединение полученных моделей погрешности мемристивного устройства и веса синапса нейрона позволяет связать напрямую параметры сигнала и выходное значение веса что является важным при проектировании ИНСМ и позволяет экономить время, за счет сокращения промежуточных операций связанных с вычислением сопротивлений.

3) Разработан алгоритм оценки ФК ИНСМ, который учитывает взаимосвязь между параметрами сигнала задания сопротивления мемристивного устройства и итоговым значением ФК ИНСМ. Данный алгоритм в отличие от других учитывает ограничения максимально допустимых напряжений, подаваемых на вход кроссбар-массива мемристивых устройств, а также использует масштабирование выходов нейронов, для повышения оценки ФК ИНСМ.

4) Предложенные модели и алгоритмы были апробированы на тестовом примере, в котором аппаратная часть смоделирована в программе LTSpice. В результате анализа моделей были выявлено, что малые значение сопротивления мемристивного устройства или весового коэффициента синапса нейрона приводят к большим значения погрешностей интерполяции. Так как задание сопротивлений мемристивного устройства является нелинейным, то на некоторых участках, при малых сопротивлениях погрешность модели возрастает, так как мемристивное устройство ведет себя нелинейно что требует большего числа точек данных для повышения качества интерполяции на данном диапазоне. Анализ данных полученных в результате применения данных моделей показал, что даже относительно небольшая погрешность мемристивного устройства может приводить к разбросу значений весового коэффициента в разы превосходящему погрешность устройств из которых оно было сделано.

Данные результаты получены автором лично и были частично опубликованы в соавторстве (вклад автора более 50 %) в [100–103,106–115].

3 Программно-аппаратный комплекс для оценки функциональной корректности искусственных нейронных сетей на базе мемристоров

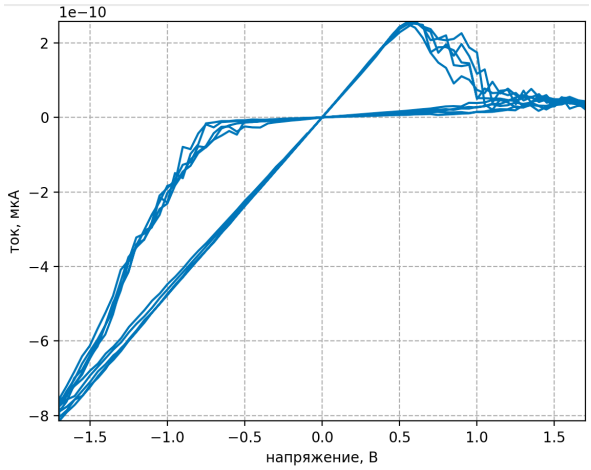
3.1 Описание исследуемых металл-оксидных мемристивных устройств

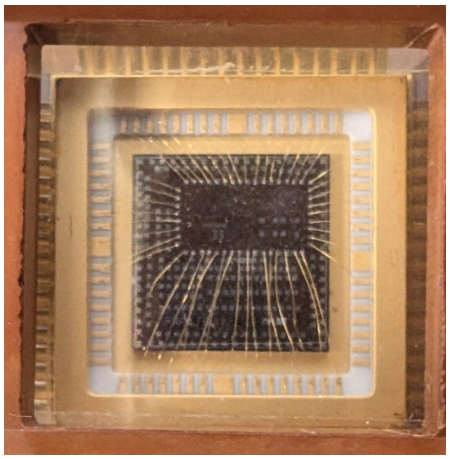
Для практического применения разработанных в рамках данного диссертационного исследования методологических средств было решено разработать специализированный программно-аппаратный комплекс (ПАК). В первую очередь данный ПАК должен содержать мемристивные устройства, для чего лабораторией мемристивной наноэлектроники НОЦ ФТНС ННГУ им. Н.И. Лобачевского были предоставлены мемристивные устройства, реализованные на основе тонкопленочной структуры «металл-диэлектрик-металл», конструктивные варианты и технологические приемы изготовления которой совместимы с современным КМОП-процессом.

В качестве диэлектрического материала в данных устройствах используется стабилизированный иттрием диоксид циркония $ZrO_2(Y)$. Это оксид переходного металла с преобладающим ионным характером химической связи (известный как ионный электролит), в котором легирующая добавка оксида иттрия (12 мол. %) стабилизирует кубическую фазу оксида металла и задает определенную концентрацию кислородных вакансий, которые играют определяющую роль в резистивном переключении – образовании и локальном разрушении проводящих каналов (филаментов) в пленке оксида.

Исследуемые мемристивные устройства имеют структуру $Au/Ta/ZrO_2(Y)/Pt/Ti$. Мемристивные устройства демонстрируют воспроизводимое биполярное резистивное переключение как в непрерывном, так и импульсном режимах при напряжениях в диапазоне 0,5-1,5 В (рисунок 3.1 а). Переход SET (из СВС в СНС) соответствует формированию филаментов в слое оксида, а переход RESET (из СНС в СВС) связан с окислением филаментов вблизи границы раздела

с активным электродом. Оба состояния характеризуются нелинейной ВАХ и низким разбросом значений сопротивления.





а) б)

Рисунок 3.1 – Мемристивные устройства: а) вольтамперные характеристики; б) фотография корпусированной микросхемы

Мемристивные устройства выполнены в форме тестового кристалла. Площадь кристалла – 10 × 10 мм^2, что обеспечивает его монтаж в металлокерамический корпус марки 5134.64-6 (рисунок 3.1 б).

Для исследования использовались мемристивные устройства, конструктивно выполненные в виде 64-битного слова – массива из 64-х мемристивных микроустройств 20 × 20 мкм^2 с одной общей шиной для нижнего электрода (с 4-мя возможными контактами к общей шине).

Корпусированные мемристивные устройства были размещены в контактные устройства УК64-4С с 64 выходными контактами, для их защиты от негативных внешних воздействий и повышения удобства работы.

3.2 Описание аппаратной части программно-аппаратного комплекса

Для работы с мемристивными устройствами в рамках диссертационного исследования был разработан ПАК (рисунок 3.2), который, с одной стороны, позволяет собирать экспериментальные данные для моделирования в соответствии

с предложенными методами, и, с другой стороны, выполнять инференс ИНСМ непосредственно на реальных устройствах.

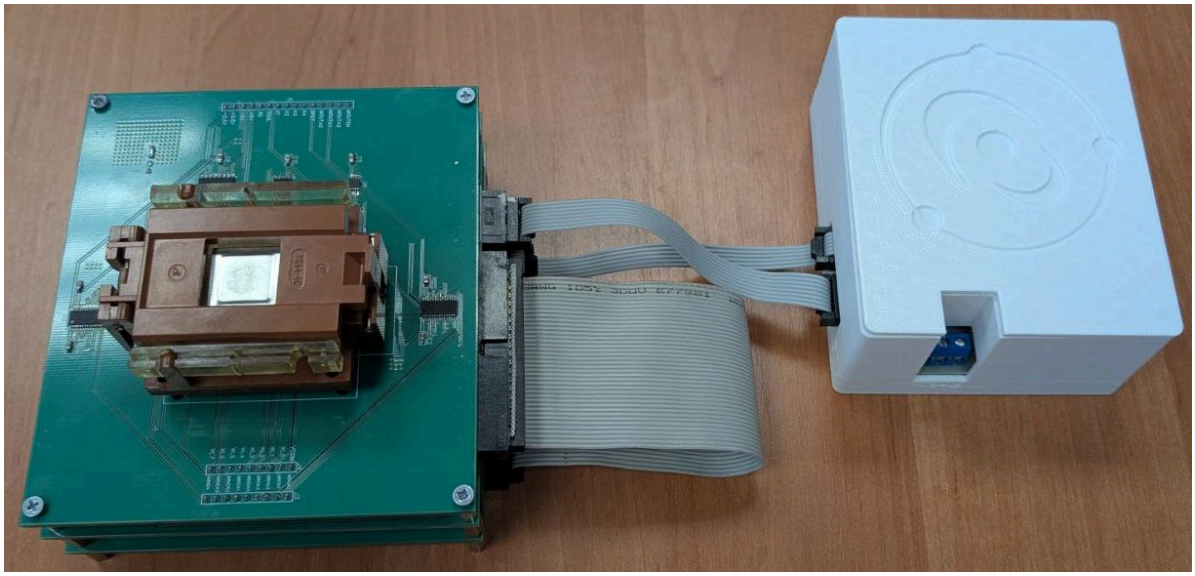


Рисунок 3.2 – Аппаратная часть ПАК

Аппаратная часть разработанного ПАК реализована в виде стека из трех плат в мезонинном соединении, а именно платы микроконтроллера (МК), платы формирования и регистрации сигналов (сигнальной платы) и коммутационной платы (для выбора нужной ячейки в кроссбар-массиве). Функциональная схема аппаратной части ПАК представлена на рисунке 3.3.

Платы ПАК выполнены на основе двухсторонней печатной платы с поверхностным монтажом компонентов. Питание аппаратной части осуществляется непосредственно от блока питания (коробка белого цвета на рисунок 3.2), который можно питать от лабораторных источников (±12 В).

Подключение аппаратной части к компьютеру осуществляется по средством USB интерфейса расположенного на плате микроконтроллера.

Корпусированные мемристивные устройства, размещённые в контактирующие устройства УК64-4С с 64 выходными контактами, могут быть подключены к ПАК с помощью разъёмов, расположенных на коммутационной плате.

Плата с выбранным мемристивным устройством может работать в двух основных режимах – RESET и SET. В режиме RESET генератор подает напряжение на верхний электрод подключенного мемристивного устройства, при этом

выполняется перевод мемристивного устройства из низкоомного состояния в высокоомное. В режиме SET генератор подает напряжение на нижний электрод подключенного мемристивного устройства, при этом выполняется перевод мемристивного устройства из высокоомного состояния в низкоомное.

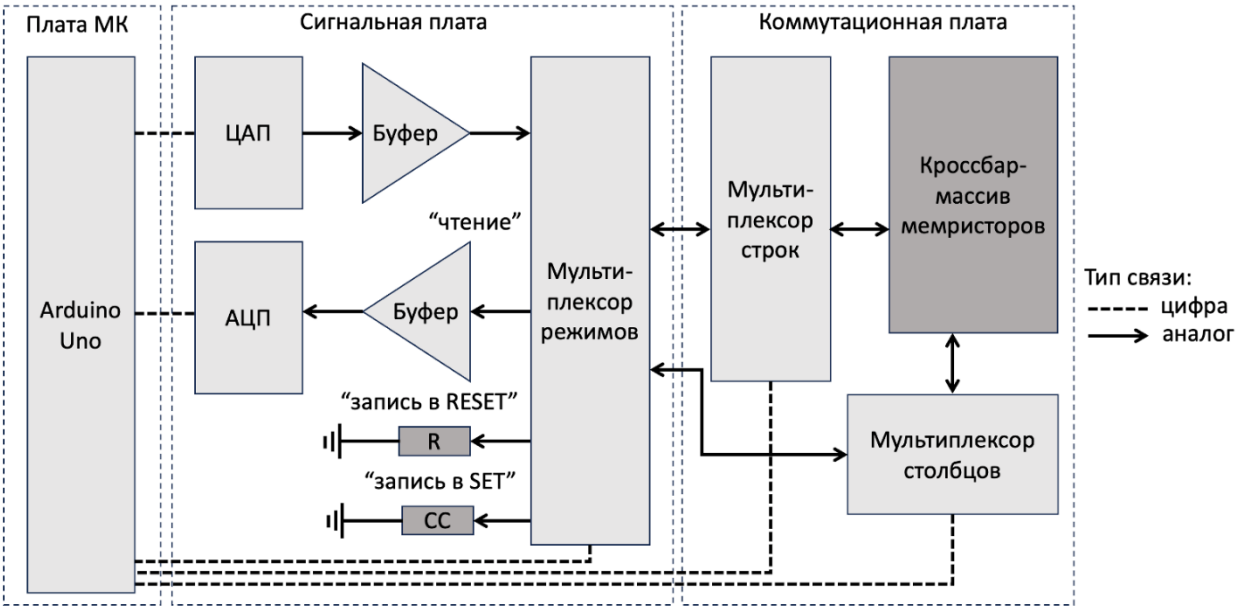


Рисунок 3.3 – Функциональная схема аппаратной части разработанного ПАК

Любое обращение к плате отрабатывается путем формирования двух импульсов, а именно «записи» (может подаваться как в RESET, так и в SET, при этом АЦП не задействуется) и «чтения» (подается в RESET и на АЦП). Используя такой подход к генерации сигналов, можно задать произвольную импульсную последовательность путем комбинации «записи» и «чтения» и подать ее на МУ. Для импульсной последовательности можно задавать различные значения параметров сигналов (в соответствии с рисунком 2), тем самым реализуя планы экспериментов с разным количеством факторов (рисунок 3.4).

Соответственно, если нужно измерить сопротивление мемристивного устройства без изменения его сопротивления, то на плату посылается директива «записи», в которой отключаются импульсы «записи» (амплитуда записи 0 В), а так как обращение состоит из двух частей, то после него всегда будет идти чтение.

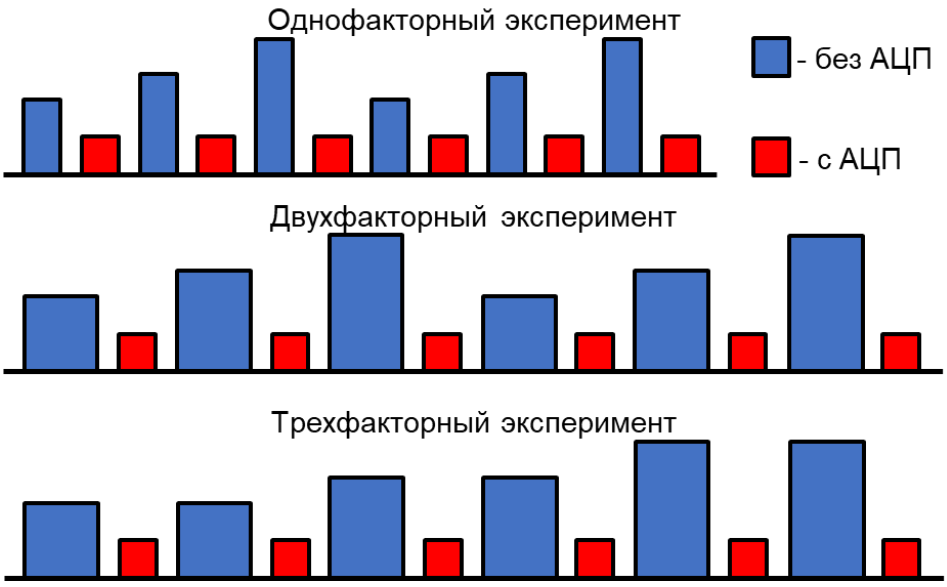


Рисунок 3.4 – Примеры реализации планов экспериментов с помощью ПАК

Представленные на рисунке 3.4 сигналы «чтения» и «записи» имеют следующие характеристики. Минимальное возможное время подачи сигнал «чтения» (красные столбцы), которое можно установить при работе с ПАК составляет ~1,2 мс. При этом во время его подачи выполняется считывание сопротивления десять раз и для полученных данных находится среднее значение. Это нужно для повышения точности измерения, и снижения влияния шумов на конечный результат. Амплитуда каждого импульса «чтения» составляет 300 мВ (рисунок 3.5). Параметры сигнала чтения:

– нарастание сигнала ~2 мкс;

– удержание сигнала ~1196 мкс;

– спад сигнала ~2 мкс.

Ограничение максимальной амплитуды импульса «чтения» обусловлено особенностями работы с мемристивными устройства на основе оксида циркония, которые были использованы в данной работе, так как именно такая амплитуда позволяет выполнять считывание сопротивления, не изменяя резистивного состояния устройств.

В свою очередь, минимальное возможное время подачи сигнала «записи» (синие столбцы) составляет 12 мкс (рисунок 3.6) из которых:

– нарастание сигнала ~2 мкс;

– удержание сигнала ~8 мкс;

– спад сигнала ~2 мкс.

Амплитуда каждого импульса «записи» может варьироваться от 0 до 3 В с разрядностью 12 бит на диапазон от 0 до 5 В.

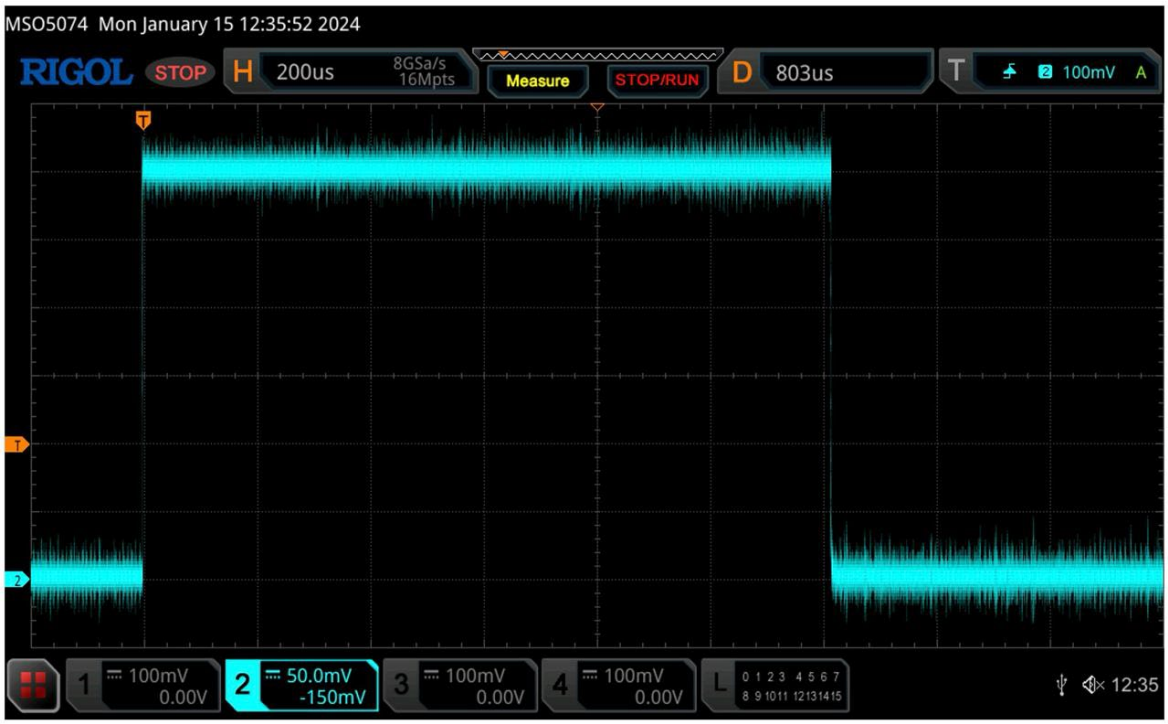


Рисунок 3.5 – Экранные изображения с осциллографа, демонстрирующие характеристики импульса чтения

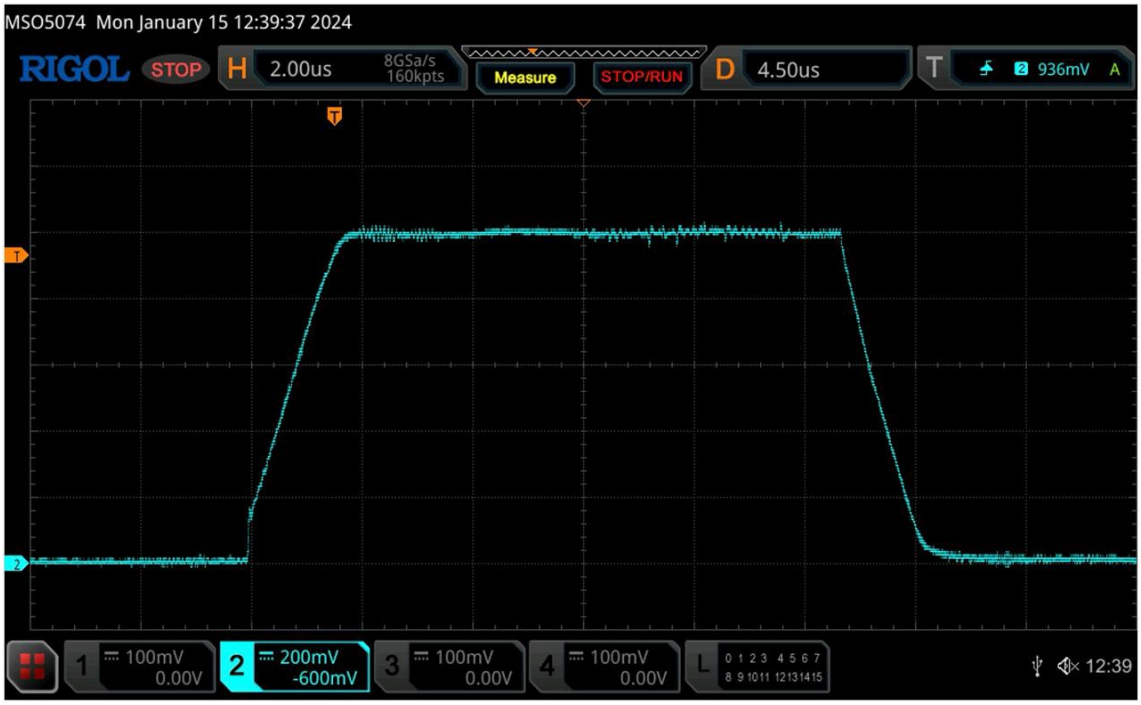


Рисунок 3.6 – Экранные изображения с осциллографа, демонстрирующие характеристики импульса «записи»

Одну посылку сигнала обращения к плате, состоящего из импульса «чтения» и «записи», можно выполнять не раньше, чем раз в 55 мс (рисунок 3.7).

ПАК имеет следующие характеристики:

– количество каналов ЦАП – 1, разрядность 12 бит;

– количество каналов АЦП – 1, разрядность 12 бит;

– генератор сигнала -3…3 В, максимальный ток до 50 мА, минимальная длительность импульса 12 мкс;

– управление ключами селекторов 0/3,3 В;

– размер подключаемого кроссбар-массива до 32x8 мемристивных устройств.

Основные особенности ПАК:

– модульная система - модуль управления, сигнальный и коммутационный;

– работа с активными кроссбар массивами МУ в архитектуре 1T1R;

– питание аппаратной части от лабораторных источников питания (±12 В).

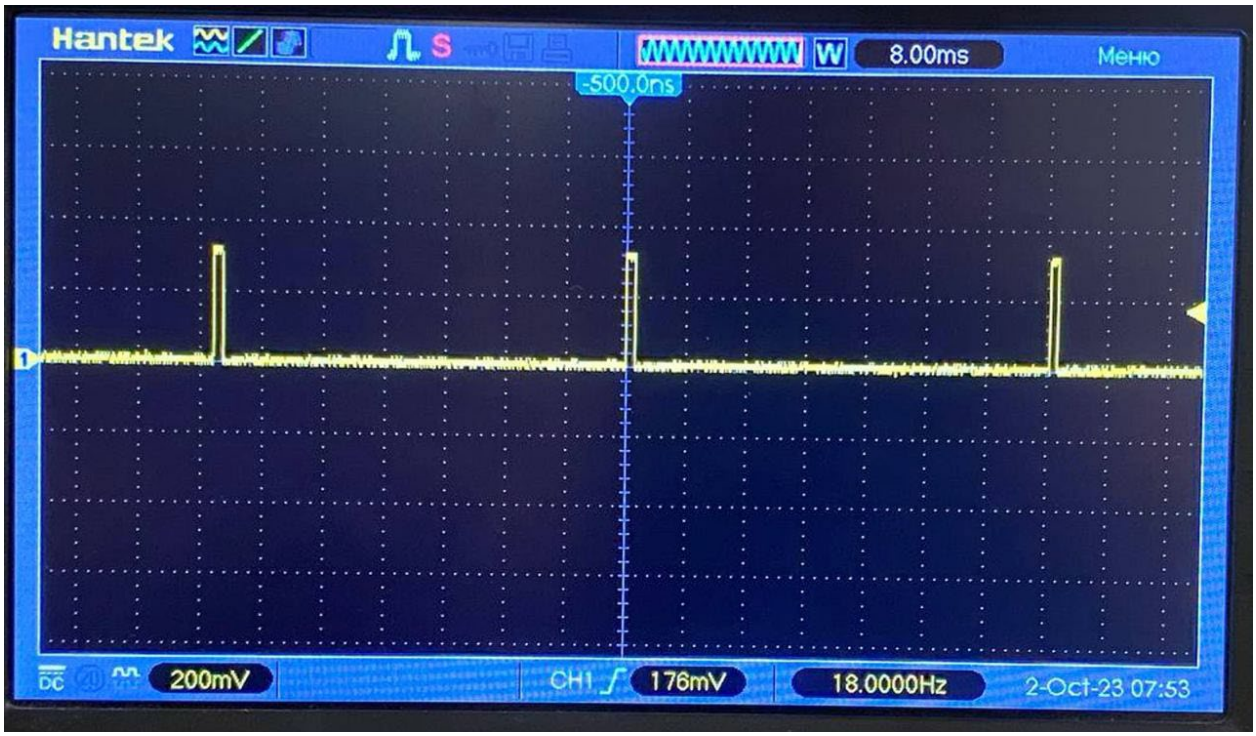


Рисунок 3.7 – Экранные изображения с осциллографа, демонстрирующие отправку сигнала «чтения» с максимально доступной для данного устройства частотой.

3.3 Описание программной части программно-аппаратного комплекса

Пользовательское программное обеспечение (ПО) ПАК было реализовано на языке программирования Python. Для удобства работы с ПО также был реализован графический интерфейс пользователя (рисунок 3.8).

Обмен данными между аппаратной частью и компьютером с установленным ПО, соединенных USB интерфейсом, осуществляется по COM-порту. Запрос к плате через COM-порт стандартизирован и состоит из следующих обязательных элементов:

– Номер команды Arduino. Данный параметр представляется в виде целого числа и определят, то, что плата должна сделать. В настоящее время плата может выполнять только три команды, а именно генерировать импульсы «чтения», «записи» и комбинация из обоих импульсов, причем сначала идет «запись», а потом «чтение».

– Напряжение, выдаваемое ЦАП. Данный параметр подставляется в виде целого числа, которое является количеством отсчетов соответствующие требуемому напряжению. Преобразование значений напряжения в отчеты осуществляется по формуле (3.1).

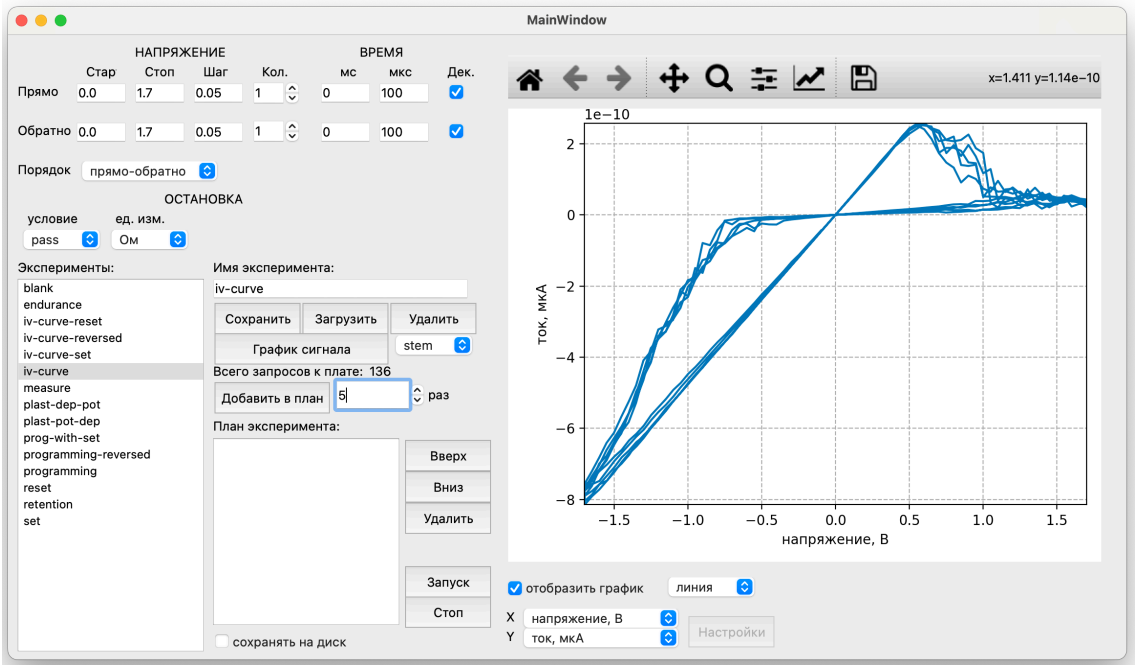


Рисунок 3.8 – Основное окно ПО

$$DAC = \frac{V_{DAC} \times 2^{B_{DAC}} - 1}{V_{RefDAC}} \quad (3.1)$$

где DAC – отсчеты, подаваемые в ЦАП для установки напряжения;

V_{DAC} – желаемое напряжение, которое должен выдавать ЦАП;

B_{DAC} – разрядность ЦАП;

V_{RefDAC} – опорное напряжение ЦАП.

– Время подачи сигнала в миллисекундах. Данный параметр представляется в виде целого числа, которое отвечает за то, сколько миллисекунд будет подаваться сигнал записи.

– Время подачи сигнала в микросекундах. Данный параметр подставляется в виде целого числа, которое отвечает за то, сколько микросекунд будет подаваться сигнал записи. Данное значение суммируется с миллисекундами на плате.

– Знак напряжения. Данный параметр подставляется в виде целого числа и отвечает за переключение режима работы платы между RESET и SET.

– Идентификатор запроса. Данный параметр подставляется в виде целого числа, которое отвечает за идентификацию отправленного запроса. То есть если на плату был отправлен запрос на выполнение какого-либо действия, то ответ от платы вернется с тем же идентификатором, с которым он был послан.

Разработанное ПО может быть использовано для 3 основных направлений деятельности:

– Исследование характеристик мемристивных устройств, а именно снятие ВАХ, устойчивости, синаптической пластичности и других характеристик.

– Создание модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания (модель 1) и модели зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса (модель 2) с применением разработанных алгоритмов 1 и 2 соответственно.

– Оценка ФК ИНСМ с применение разработанного алгоритма 3.

– Аппаратный инференс ИНСМ.

3.3.1 Исследование характеристик мемристивных устройств

Если рассматривать данное ПО с точки зрения исследования характеристик мемристивных устройств, то оно позволяет осуществлять следующие:

– Считывать сопротивление мемристивного устройства.

– Задавать сопротивление мемристивного устройства с требуемым допуском на значение резистивного состояния.

– Переводить мемристивное устройство в низкоомное и высокоомное состояние.

Важно также отметить, что так как получаемое с платы значение сопротивления мемристивного устройства приходит в формате отсчетов АЦП, то для его перевода в сопротивление в ПО применяется формула (3.2).

$$R_m = \frac{G \times R_l \times V_r \times 2^{B_{ADC}}}{\left(ADC \times V_{RefADC}\right)} - R_s - R_l \qquad (3.2)$$

где R_m – сопротивление мемристивного устройства;

G – коэффициент усиления;

R_l – сопротивление нагрузочного резистора;

V_r – напряжение чтения;

B_{ADC} – разрядность АЦП;

ADC – отсчеты, получаемые от АЦП;

V_{RefADC} – опорное напряжение АЦП;

R_s – сопротивление переключателей.

При необходимости пользователь может сам формировать требуемые для его исследования сигналы, которые можно подать на мемристивное устройство, используя соответствующий инструментарий, а именно:

– Выбирать направление сигнала (прямой или обратный) и его порядок.

– Изменять напряжение сигнала.

– Задавать диапазон импульсов последовательно возрастающих в определенном диапазоне сопротивлений с выбранным шагом или последовательно убывающих.

– Изменять количество импульсов и длину каждого импульса с точностью до микросекунд, с учетом аппаратных ограничений.

Такой механизм настройки сигналов, подаваемых на мемристивные устройства, позволяет добиваться различных форм и соответственно реализовывать различные эксперименты, выходящие за рамки предопределенных настроек.

Кроме того, если пользователю необходимо выполнить несколько последовательных экспериментов, отличающихся формой сигнала, он может сформировать план эксперимента, который в автоматическом режиме выполнит все добавленные в него эксперименты и сохранит полученные результаты.

Для визуализации результатов экспериментов, а также подаваемых сигналов, в правой части программы предусмотрена панель, отображающая соответствующие графики (рисунок 3.9).

При установке сопротивлений мемристивных устройств пользователь также может выбрать один из режимов остановки, таких как остановка при превышении какого-либо значения, при попадании в указанный пользователем диапазон и т.д. Соответственно для перевода значения сопротивления в отсчёты АЦП применяется формула (3.3).

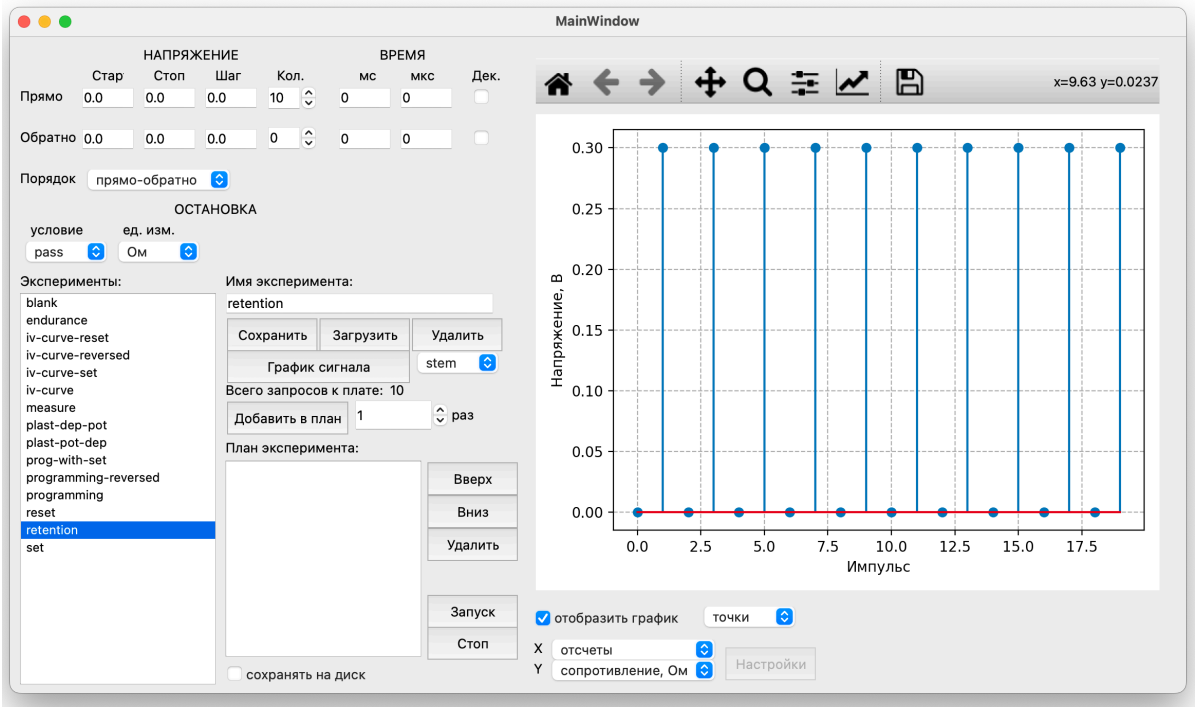


Рисунок 3.9 – Окно ПО, с визуализацией сигнала для снятия «retention»

$$ADC = \frac{G \times R_l \times V_r \times 2^{B_{ADC}}}{V_{RefADC} \times (R_s + R_l + V_{ADC})} \quad (3.3)$$

где V_{ADC} – желаемое напряжение, которое должен выдавать АЦП;

G – коэффициент усиления;

R_l – сопротивление нагрузочного резистора;

V_r – напряжение чтения;

B_{ADC} – разрядность АЦП;

ADC – отсчеты, получаемые от АЦП;

V_{RefADC} – опорное напряжение АЦП;

R_s – сопротивление переключателей.

3.3.2 Создание модели погрешности мемристивного устройства и веса синапса нейрона

Если рассматривать данное ПО с точки создания модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания, то оно позволяет создавать различные модели как на основе реальных данных, так и на основе модельных.

Используя данное ПО, пользователь может выполнить одномерную, двумерную или трехмерную интерполяцию сопротивления мемристивного устройства и его погрешности в зависимости от различных параметров сигнала задания сопротивления, таких как:

– Амплитуда импульсов;

– Длительность импульсов;

– Количество импульсов.

Кроме того, данное ПО позволяет выполнять интерполяцию обратной функции.

В качестве данных для построения моделей могут выступать не только реальные данные полученные непосредственно с мемристивных устройств, но и

данные, полученные из их моделей. Для этого в программе реализован интерфейс взаимодействия с программой LTSpice. Его суть заключается в том, что он позволяет запускать копию программы LTSpice в фоновом режиме с требуемыми параметрами модели и получать результаты моделирования.

Соответственно для получения модельных данных была создана схема записи сопротивления мемристивного устройства представленная на рисунке 2.5.

Для создания модели 1 сопротивление МУ пересчитывается из цифрового кода АЦП по формуле

$$R_m = \frac{K_g \times R_l \times V_r \times 2^{14}}{DOC \times V_{ra}} - R_s - R_l, \quad (3.4)$$

где R_m – сопротивление мемристора;

DOC – цифровой код АЦП;

K_g – коэффициент усиления измерителя (11);

R_l – сопротивление нагрузки (3000 Ом);

V_r – напряжение чтения (0,3 В);

V_{ra} – опорное напряжение АЦП (5 В);

R_s – сопротивления ключей и других элементов цепи (10 Ом).

Таким образом для того, чтобы изменить модель пользователь может либо подстроить имеющую модель под параметры своих мемристивных устройств либо, заменить на ту модель, которая нужна, настроив требуемы параметры как в модели схемы, так и в самом ПО.

Моделирование работы схемы в LTSpice может занимать достаточно большое количество времени при выполнении больших планов экспериментов содержащих десятки экспериментов с тысячами повторений. Поэтому, для ускорения, программа распараллеливает выполнение экспериментов, создавая несколько фоновых копий программы LTSpice, которые параллельно осуществляют моделирование. Такой подход позволяет уменьшить время моделирования при этом пользователь сам может выбирать сколько процессов он хочет для этого использовать.

План эксперимента для получения данных о задании сопротивления мемристивного устройства как для модели, так и для реальных мемристивных устройств можно собрать, используя механизм настройки сигналов, который описан в пункте 3.3.1.

В качестве интерполяции данных в ПО реализовано два метода – кусочно-линейная и кусочно-кубическая интерполяция. Полученные в результате интерполяции модели могут быть представлены как в виде программных моделей, в которые пользователь просто подставляет данные и получает соответствующие им значение, так и в виде математических формул, которые можно затем интегрировать в другие системы моделирования и реализовывать на других языка программирования.

В данном ПО для создания модели погрешности веса была выбрана схема реализации синапса из двух и одного мемристоров, как одни из наиболее популярных.

В качестве интерполяции данных значений весов в ПО, также используются два метода – кусочно-линейная и кусочно-кубическая интерполяция. Полученные в результате интерполяции модели могут быть представлены как в виде программных моделей, в которые пользователь просто подставляет данные и получает соответствующие им значение, так и в виде математических формул, которые можно затем интегрировать в другие системы моделирования и реализовывать на других языка программирования.

Полученные таким образом модели затем могут быть использованы при компьютерном моделировании ИНСМ. Полученные модели в процессе моделирования используются для перечёта номинальных значений весовых коэффициентов с учетом погрешности мемристивных устройств.

Другим возможным применением данных моделей является использование их с точки зрения синтеза, когда берутся все веса ИНС и затем для каждого из них определяются с помощью модели параметры сигнала задания сопротивления, которые позволяют достичь требуемых значений. Затем на основе полученных данных строится план эксперимента, в котором собираются данные о

сопротивлениях мемристивных устройств и затем полученные данные используются для моделирования ИНСМ.

3.3.3 Оценка функциональной корректности искусственной нейронной сети на базе мемристоров

Для оценки ФК ИНСМ в ПО предусмотрены следующие возможности:

– Загрузка различных моделей ИНС, созданных с применением библиотеки Keras (TensorFlow);

– Изменение количества экспериментов.

– Изменение масштабирующего коэффициента *K*.

– Изменение ограничения максимально допустимых напряжений, подаваемых на вход кроссбара мемристивных устройств *T*.

– Расчет метрики ФК ИНСМ.

3.3.4 Аппаратный инференс искусственной нейронной сети на базе мемристоров

Если рассматривать данное ПО с точки зрения аппаратного инференса ИНСМ, то оно позволяет осуществлять следующие:

– Выполнять маппирование весовых коэффициентов синапсов нейронов.

– Выполнять масштабирование входных данных и весовых коэффициентов синапсов нейронов.

– Эмулировать отрицательные значения весовых коэффициентов для схем синапса из одного мемристора.

– Оценивать метрику ФК ИНСМ при аппаратной реализации.

3.4 Выводы по главе

1) Для практического применения разработанных в рамках данного диссертационного исследования методологических, аппаратных и программных средств выбраны мемристивные устройства, имеющие структуру Au/Ta/ZrO_2(Y)/Pt/Ti. Мемристивные устройства демонстрируют воспроизводимое биполярное резистивное переключение как в непрерывном, так и импульсном режимах при напряжениях в диапазоне 0,5-1,5. Процесс SET и RESET характеризуются нелинейной вольтамперной характеристикой и низким разбросом значений сопротивления. Выбранные устройства можно использовать для создания ИНСМ.

2) Разработан ПАК для моделирования ИНСМ. Данный пак позволяет формировать сигналы различной амплитуды и длительности с помощью 12 битного ЦАП для организации широкого спектра возможностей по взаимодействию с мемристивными устройствами. При этом для считывания данных используется 12 битный АЦП. Мемристивные устройства, размещённые в контактные устройства УК64-4С, могут быть подключены к ПАК с помощью разъёмов, расположенных на коммутационной плате. Подключение к компьютеру аппаратной части ПАК осуществляется по USB интерфейсу, а обмен данными между аппаратной и программной частью выполняется с помощью стандартизированных команд, пересылаемых по COM-порту. Разработка ПАК, как испытательного стенда, является обязательным этапом подготовительных работ алгоритма оценки качества систем ИИ в соответствии с ГОСТ Р 59898-2021.

3) Программная часть ПАК реализована в виде кроссплатформенного приложения с графическим интерфейсом пользователя. Программная часть предоставляет пользователю широкий инструментарий по исследования мемристивных устройств и проведению многофакторных экспериментов. Также в программе предусмотрен функционал для создания модели погрешности мемристивных устройств, как на основе реальных данных, полученных в

результате их записи, так и на основе модельных дынных сгенерированный в LTSpice по алгоритму 1 и моделей погрешности весов синапсов нейронов по алгоритму 2. В дополнение в программной части предусмотрена возможность оценки ФК ИНСМ с применением модели 2 и алгоритма 3 и аппаратный инференс на мемристивных устройствах.

Данные результаты получены автором лично и были частично опубликованы в соавторстве (вклад автора более 50 %) в [100,101,109–115].

4 Практическое применение разработанных моделей и алгоритмов, аппаратных и программных средств

4.1 Построение модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания

Для построения модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания в первую очередь был определен диапазон сопротивлений используемых мемристивных устройств с помощью подачи одиночных прямоугольных импульсов в режимах RESET и SET с применением разработанного ПАК. В результате было определено, что в среднем R_{min} = 8500 Ом, R_{max} = 75000 Ом, а задание минимального низкоомного состояния можно осуществить одним импульсом с амплитудой -1,7 В и длительностью 100 мкс.

В соответствии с первым шагом алгоритма 1 были заданы исходные данные по параметрам сигнала записи, которые вводятся в программу. В данном случае амплитуда сигнала имела 3 уровня – 0,8, 1,1, и 1,7 В, количество импульсов также имело 3 уровня – 1, 10 и 19 штук, а длительность была фиксированной и составляла 100 мкс. Такой выбор факторов был сделан в результате предварительного анализа используемых мемристивных устройств с помощью ПАК – было выявлено, что наиболее сильное влияние на процесс изменения сопротивления оказывает амплитуда сигнала, при этом и увеличение количества, и увеличение длительности импульса позволяют менять сопротивление, но только до определенного предела (R_{max}), и комбинация этих двух параметров не позволяет его преодолеть. Значения факторов были выбраны таким образом, что они позволяют практически полностью покрыть диапазон сопротивлений. Это будет видно далее из результатов экспериментов.

В соответствии с шагом 2 алгоритма 1 был сформирован план двухфакторного эксперимента, в котором факторами являются амплитуда импульса и количество импульсов (таблица 4.1).

Таблица 4.1 – План эксперимента для задания сопротивления мемристивного устройства

Номер эксперимента	**Амплитуда импульса, В**	**Количество импульсов, шт.**
1	0,8	1
2	0,8	10
3	0,8	19
4	1,1	1
5	1,1	10
6	1,1	19
7	1,7	1
8	1,7	10
9	1,7	19

Затем в соответствие с шагом 3 и 4 выполнялись циклы по экспериментам и по параллельным опытам. Количество параллельных опытов было 1000 штук для каждого уровня. В циклах выполнялась подача на мемристоры сигналов в соответствии с планом эксперимента (шаг 5) (таблица 4.1) и накапливались полученные значения *R* (шаг 6), которые были визуализированы на рисунке 4.1.

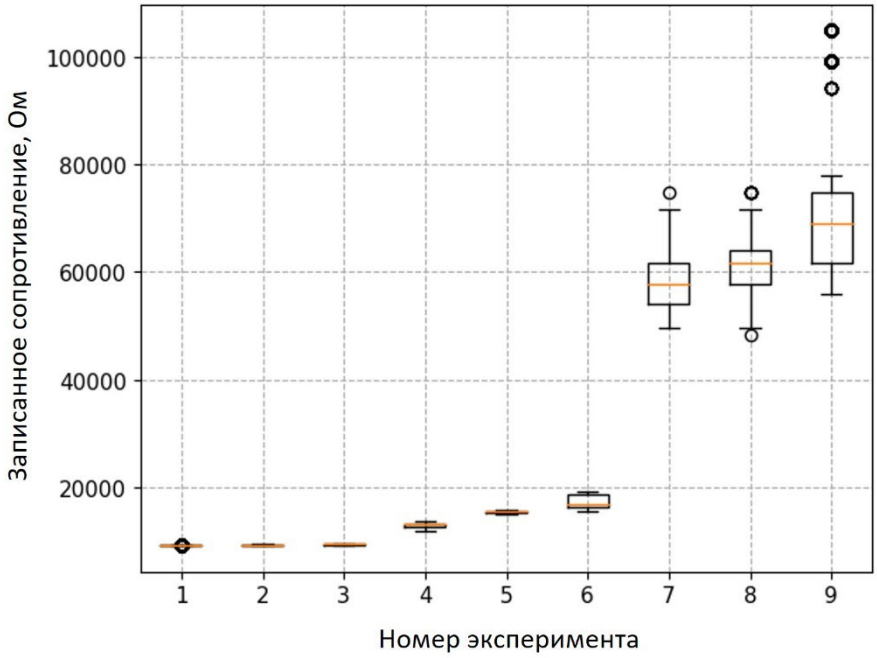


Рисунок 4.1 – Результаты накопления значений *R*

Как можно видеть из рисунка 4.1, результаты эксперимента имеют некоторое количество выбросов, которые могут негативно повлиять на модель, создаваемую на их основе, поэтому данные выбросы были выявлены с помощью межквартильного размаха и удалены. Затем в соответствие с шагом 7 была выполнена оценка МО и СКО для полученных сопротивлений (таблица 4.2).

Таблица 4.2 – Параметры закона распределения для интерполяции мемристивного устройства

Номер эксперимента	**МО сопротивления, Ом**	**СКО сопротивления, Ом**
1	9079	0
2	9201	47
3	9300	74
4	12724	492
5	15267	902
6	16972	1312
7	58642	5384
8	60709	5509
9	72225	5634

Далее в соответствии с шагом 8 и 9 алгоритма 1 была выполнена интерполяция и построена модель зависимости сопротивления *R* МУ от параметров сигналов его задания *F*. Для интерполяции был выбран метод кусочно-линейной интерполяции. Графики полученных моделей представлены на рисунке 4.2 и 4.3.

Далее была выполнена интерполяция СКО относительно МО сопротивления мемристивного устройства. Данная модель в дальнейшем будет использована для генерации данных для построения модели веса. График данной модели представлен на рисунке 4.4.

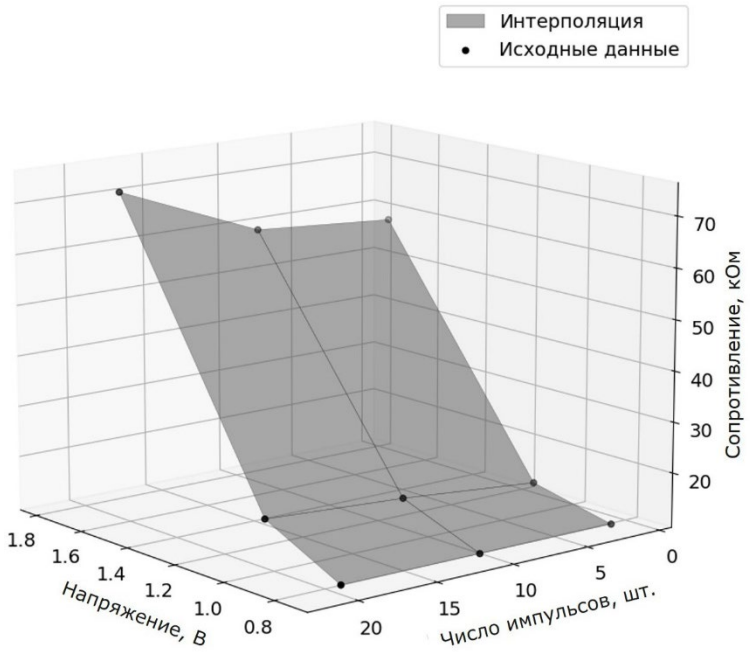


Рисунок 4.2 – Модель 1 для исследуемого кроссбар-массива: а) функция $\tilde{R} = f(F_1, \ldots, F_e)$

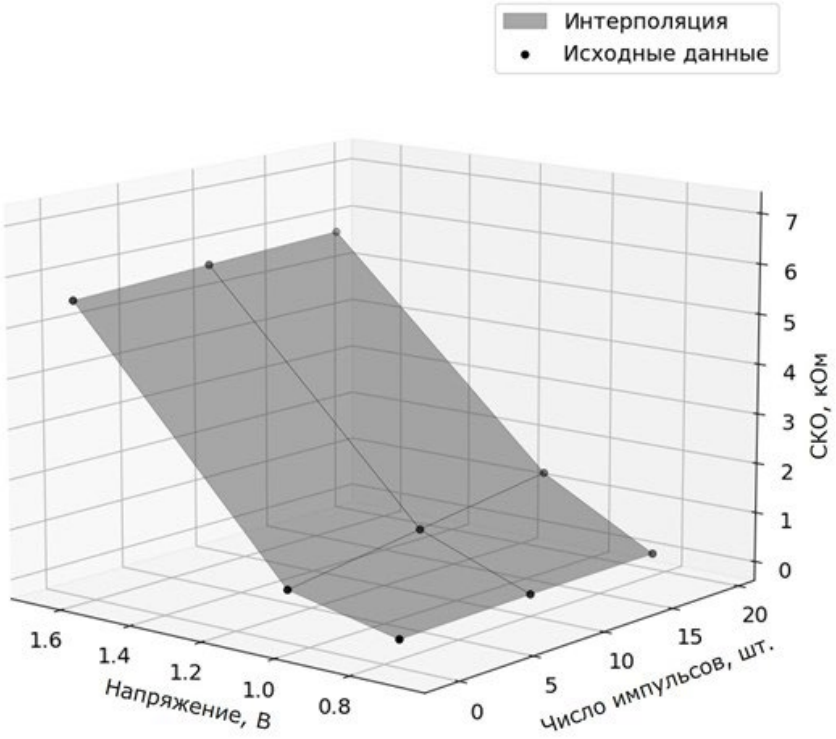


Рисунок 4.3 – Модель 1 функция $\sigma_R = k(F_1, \ldots, F_e)$

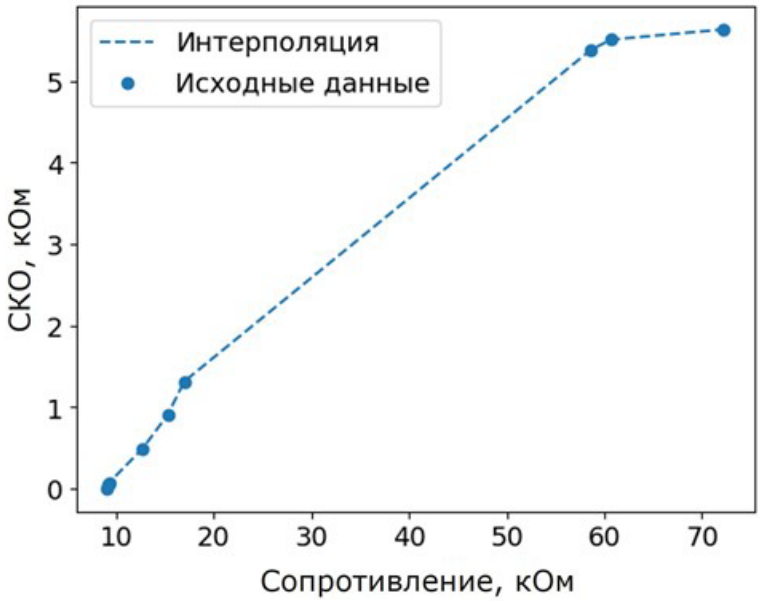


Рисунок 4.4 – Модель 1 функция $\sigma_R = s(\tilde{R})$

Для полученной модели 1 была выполнена проверка распределения полученных данных из модели и экспериментальных по тесту Колмогорова-Смирнова с помощью библиотеки SciPy для языка Python с помощью функции ks_2samp. В результате проверки было установлено, что модельные и экспериментальные данные совпадают с доверительной вероятностью 95%.

4.2 Построение модели зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса

В данном случае была использована схема синапса из одного мемристора. Такой выбор был сделан исходя из того, что данная схема требует наименьшее число мемристивных устройств, что удобно в условиях небольшого количества устройств в кроссбар-массиве 32x8, а также потому, что данная схема веса заложена в ПАК и будет далее использоваться для верификации модельных данных в эксперименте. Формула для расчета весового коэффициента при такой реализации имеет следующий вид:

$$W = \frac{R_l}{R_l + R_m}, \tag{4.1}$$

где W – вес синапса нейрона;

R_l – нагрузочное сопротивление цепи (3000 кОм);

R_m – сопротивление мемристора.

Затем в соответствии с алгоритмом 2 шагом 1 в программу вводятся уровни значений сопротивлений мемристора R, для которых будет выполняться моделирование веса. В данном случае было решено использовать теже уровни, что были получены экспериментально в предыдущем пункте, а именно 9079, 9201, 9300, 12724, 15267, 16972, 58642, 60709 и 72225 Ом.

В соответствии с шагом 2 алгоритма 2 был создан план однофакторного эксперимента для значений, которые будут подаваться в модель мемристора (таблица 4.3). В данном случае для фактора R 9 уровней.

Таблица 4.3 – План эксперимента

Номер эксперимента	**Сопротивление, Ом**
1	9079
2	9201
3	9300
4	12724
5	15267
6	16972
7	58642
8	60709
9	72225

Затем в соответствие с шагом 3 и 4 алгоритма 2 выполнялись циклы по экспериментам и по параллельным опытам. Количество параллельных опытов было 1000 штук для каждого уровня. В циклах выполнялось моделирование весового коэффициента в соответствии с планом эксперимента (шаг 5) (таблица 4.1) и накапливались полученные значения R (шаг 6), которые были визуализированы на рисунке 4.5.

В результате был накоплен набор статистических данных, представленный на рисунке 4.5.

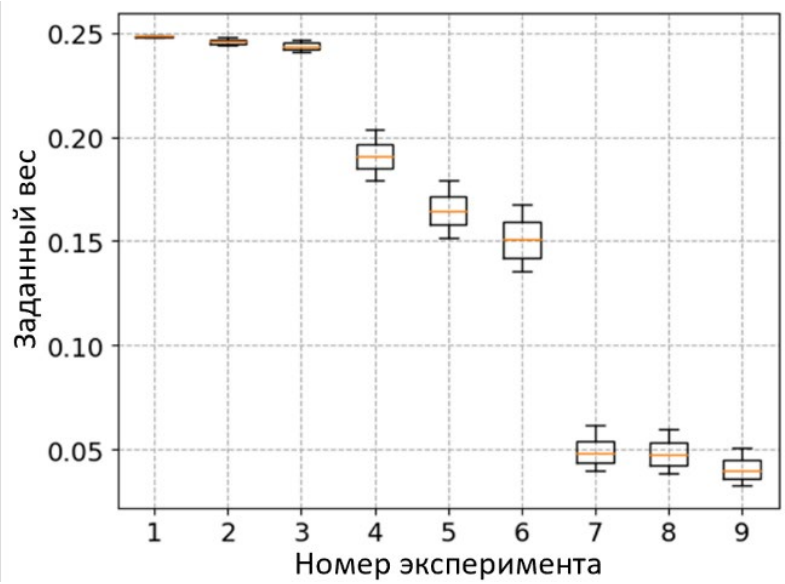


Рисунок 4.5 – Модель 2 для используемой схемы синапса нейрона (функция $\widetilde{w} = g(R)$)

Из полученного графика видно, что выбросы для полученных данных отсутствуют, поэтому дополнительного применения каких-либо методов для избавления от них не требуется. Для полученных данных было вычислено МО и СКО (шаг 7), для которых в последствии выполнялась интерполяция. Полученные значения представлены в таблице 4.4.

Таблица 4.4 – Параметры закона распределения для интерполяции веса

Номер эксперимента	**МО веса**	**СКО веса**
1	0,248	0,000
2	0,246	0,001
3	0,244	0,001
4	0,191	0,006
5	0,164	0,008
6	0,150	0,009
7	0,049	0,004
8	0,047	0,004
9	0,039	0,003

В соответствии с шагом 8 и 9 была выполнена интерполяция и построена модель зависимости веса синапса нейрона W от сопротивления мемристивного устройства R и схемы формирования веса. Для интерполяции был выбран метод кусочно-линейной интерполяции. График модели представлены на рисунке 4.6.

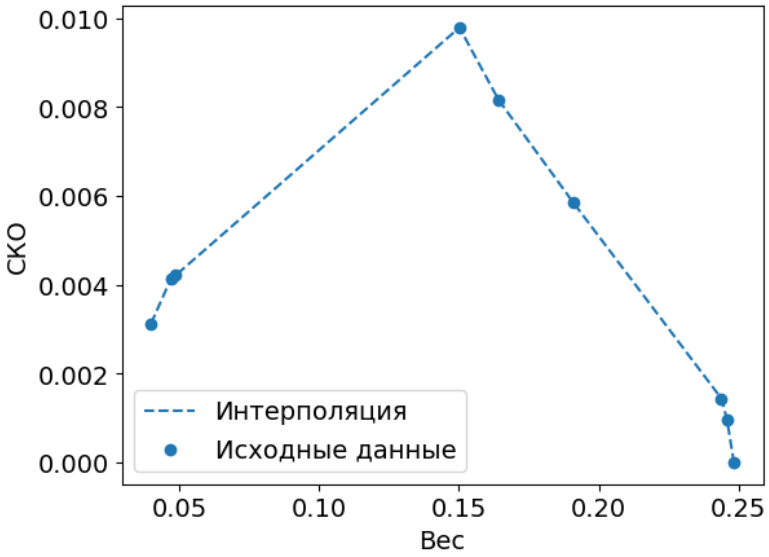


Рисунок 4.6 – Модель 2 для используемой схемы синапса нейрона (функция $\sigma_w = u(\widetilde{w})$)

Полученная модель веса в дальнейшем будет использована для оценки функциональной корректности ИНСМ с учетом вариаций сопротивлений мемристивных устройств.

Для построенной модели 2 была выполнена проверка распределения полученных данных из модели и экспериментальных по тесту Колмогорова-Смирнова с помощью библиотеки SciPy для языка Python с помощью функции ks_2samp. В результате проверки было установлено, что модельные и экспериментальные данные совпадают с доверительной вероятностью 95%.

4.3 Оценка функциональной корректности тестовых искусственных нейронных сетей на базе металл-оксидных мемристивных устройств

4.3.1 Классификация ирисов Фишера

В качестве одного из демонстрационных примеров была выбрана классическая задача машинного обучения по многоклассовой классификации на наборе данных ирисов Фишера (рисунок 4.7).

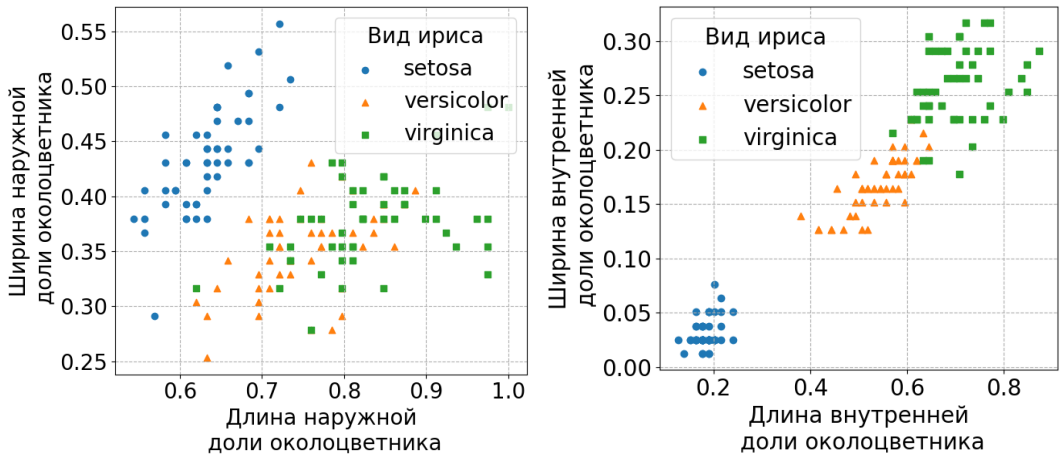
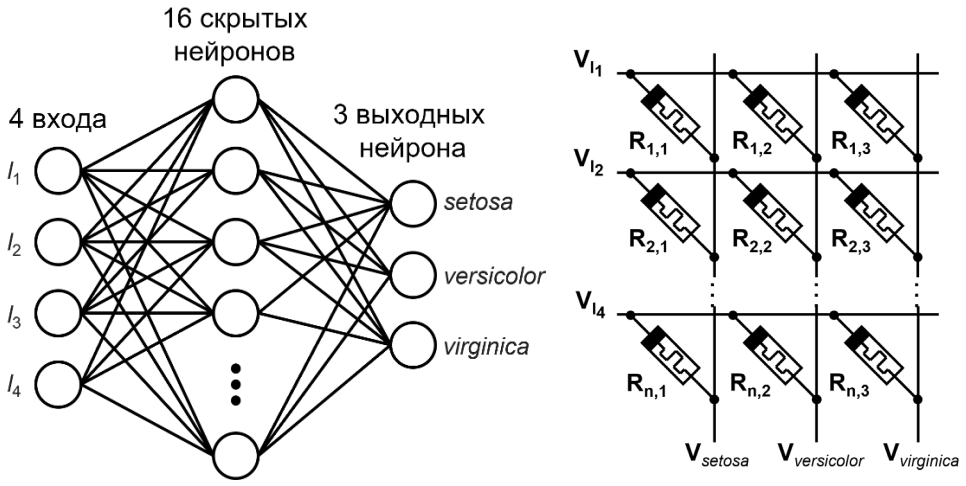


Рисунок 4.7 – Визуализация данных «Ирисы Фишера»

Этот набор данных содержит признаки для 150 экземпляров ирисов, по 50 для каждого отдельного вида – Ирис щетинистый (Iris setosa), Ирис виргинский (Iris virginica) и Ирис разноцветный (Iris versicolor). Для каждого экземпляра класса в наборе данных содержится по 4 признака, а именно длина и ширина наружной доли околоцветника, а также длина и ширина внутренней доли околоцветника. Задачей для ИНСМ в данном случае является классификация вида конкретного экземпляра ириса по 4 признакам на 3 класса.

Для решения поставленной задачи была создана искусственная нейронная сеть в архитектуре многослойного персептрона. По структуре данная ИНС имеет 4 входных нейрона, что соответствует количеству признаков каждого примера в наборе данных, 16 нейронов в скрытом слое с функцией активации ReLU каждый, а также 3 нейрона в входном слое, что соответствует количеству классов в наборе данных, с функцией активации SoftMax.

Обучение ИНС осуществлялось методом стохастического градиентного спуска, основанного на адаптивной оценке моментов первого и второго порядка. Потери в процессе обучения оценивались через расчет перекрестной энтропии между исходными метками и предсказаниями ИНС.



а) б)

Рисунок 4.8 – ИНС для классификации ирисов Фишера: а) схематичная структура ИНС; б) реализация ИНС на базе кроссбар-массива мемристивных устройств

В соответствии с ГОСТ Р 59898-2021 метрикой оценки ФК в задачах классификации может являться доля правильных исходов (A), рассчитываемая по формуле

$$A = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4.2}$$

где TP – количество истинно положительных исходов;

TN – количество истинно отрицательных исходов;

FP – количество ложно положительных исходов;

FN – количество ложно отрицательных исходов.

Перед обучением описанной выше ИНС исходные признаки были нормализованы в диапазоне от 0 до 1, а целевые значения классов преобразованы с помощью унитарного кода (one-hot encoding).

Для обучения ИНС использовались 70% примеров из набора данных, содержащие все виды классов примерно в одинаковых пропорциях, а для тестирования оставшиеся 30%. В итоге доля правильных исходов для программно реализованной модели (A_s) равна 1 для тестовых данных.

Далее было выполнено квантование весов полученной ИНС методом k-ближайших соседей для оптимизации количества используемых мемристоров на сеть и ускорения её работы. В результате в качестве оптимального значения было решено использовать 4 уникальных по модулю значений весов (4 положительных

и 4 отрицательных). Так как в качестве архитектуры синапсов была выбрана реализация синапса из 1 мемристора (пункт 3.2), а для проверки соответствия эксперименту будет использован ПАК, в котором производится поэлементное умножение, соответственно для реализации полученной квантированной сети потребуется использовать 4 мемристивных устройства. При этом, так как данная архитектура синапсов не позволяет реализовывать отрицательные весовые коэффициенты, то отрицательные значения при аппаратной реализации будут эмулироваться программно.

Далее с применением полученной модели 2 (пункт 4.2) и алгоритма 3 выполнена оценка ФК модели ИНСМ. Количество экспериментов в данном случае было равно 1000, $K = 1$, $T = 0{,}3$. В результате такого исследования было выяснено, что доля правильных исходов ИНСМ, полученная в результате компьютерного моделирования (A_M) при такой реализации, также составляет 1.

Далее для проверки качества результатов моделирования данная ИНСМ была реализована аппаратно с помощью разработанного ПАК. Для этого было выполнено маппирование весовых коэффициентов, которое осуществлялось следующим образом:

– Квантованная модель ИНСМ была загружена в ПАК.

– Из программной модели были извлечены уникальные значения весовых коэффициентов.

– Затем по формуле (4.3) было определено требуемое сопротивление для каждого веса.

$$R_m = \frac{R_l - W \cdot R_l}{R_l}. \tag{4.3}$$

– Для полученных сопротивлений с помощью модели 1 определялась амплитуда и длительность импульса.

– Из матрицы кроссбара выбирались в ручном режиме мемристивные устройства.

– Мемристоры переводились в минимальное низкоомное состояние.

– После сброса выполнялась запись требуемого сопротивления с помощью определенных ранее параметров сигнала.

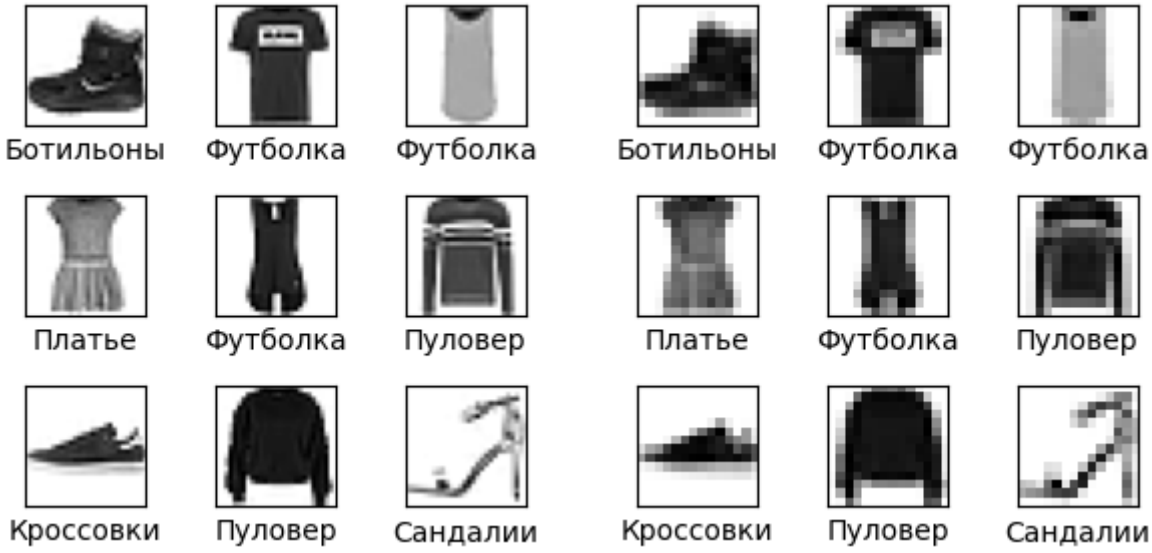
После прогона аппаратно реализованной ИНСМ доля правильных исходов (A_H) на тестовой выборке равна 1, что полностью соответствует результатам моделирования.

4.3.2 Классификация изображений элементов гардероба

В качестве еще одного демонстрационного примера была выбрана классическая задача машинного обучения по многоклассовой классификации изображений на наборе данных Fashion-MNIST.

Этот набор данных содержит 70000 изображений элементов гардероба, разделенные на 10 классов, а именно: футболка, брюки, пуловер, платье, пальто, сандалии, рубашка, кроссовки, сумка, ботильоны. Каждый экземпляр класса в наборе данных представлен в виде изображения в оттенках серого размером 28 на 28 пикселей (рисунок 4.9 а). Для сокращения числа параметров ИНС исходные изображения были масштабированы до 14 на 14 пикселей (рисунок 4.9 б). ИНСМ в данном случае необходимо по одноканальным изображениям в оттенках серого размером 14 на 14 пикселей выполнить классификацию на 10 классов.

Для решения поставленной задачи применялась сверточная нейронная сеть. По структуре данная ИНС имеет один сверточный слой с 4 ядрами свертки размерностью 3 на 3 и функцией активации ReLU, а также 10 выходных нейронов в полносвязном слое без функции активации.



а) б)

Рисунок 4.9 – Визуализация набора данных: а) исходные изображения размером 28 на 28 пикселей; б) с масштабированные исходные изображения до 14 на 14 пикселей

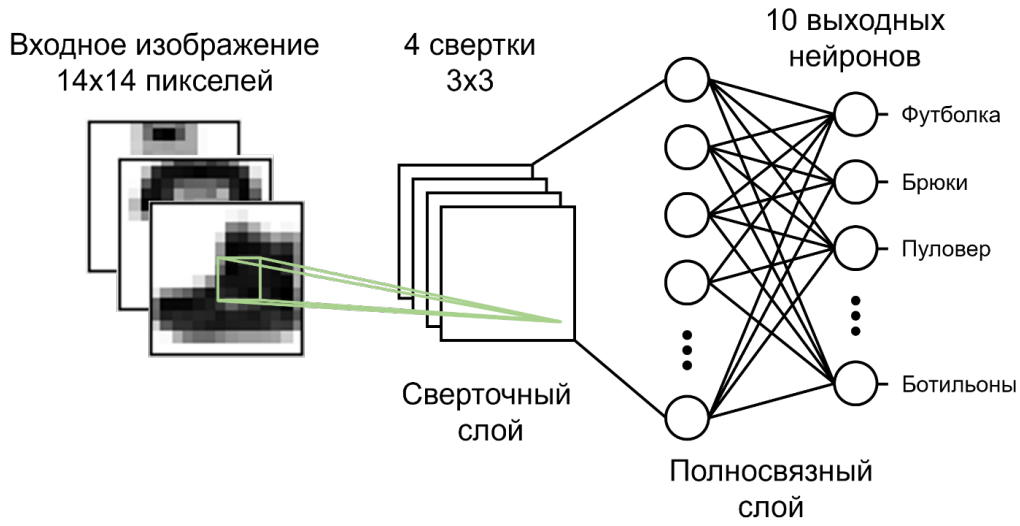


Рисунок 4.10 – ИНС для классификации элементов гардероба на наборе данных Fashion MNIST

Обучение ИНС осуществлялось методом стохастического градиентного спуска, основанного на адаптивной оценке моментов первого и второго порядка. Потери в процессе обучения оценивались через расчет перекрестной энтропии между исходными метками и прогнозами ИНС. В качестве метрики оценки ФК была выбрана доля правильных исходов (A), которая вычислялась по формуле (4.2).

Перед обучением описанной выше ИНС исходные признаки были нормализованы в диапазоне от 0 до 1, а целевые значения классов преобразованы с помощью унитарного кода.

Для обучения ИНС использовались 60000 примеров из набора данных, содержащие все виды классов примерно в одинаковых пропорциях, а для тестирования оставшиеся. В итоге доля правильных исходов для программно реализованной модели (A_s) равна 0,9 для тестовых данных.

Далее было выполнено квантование весов полученной ИНС методом k-ближайших соседей для оптимизации количества используемых мемристоров на сеть и ускорения её работы. В результате в качестве оптимального значения было решено использовать 8 уникальных по модулю значений весов (8 положительных и 8 отрицательных). В результате квантования значение доли правильных исходов для программной модели ИНС снизилось до 0,867. Для реализации полученной квантованной сети потребуется использовать 8 мемристивных устройств.

С применением полученной модели 2 (пункт 4.2) и алгоритма 3 выполнена оценка ФК модели ИНСМ. Количество экспериментов в данном случае было равно 1000, $K = 1$, $T = 0{,}3$. В результате такого исследования было выяснено, что доля правильных исходов ИНСМ, полученная в результате компьютерного моделирования (A_M) при такой реализации, составляет 0,766.

Для проверки качества модели данная ИНСМ была реализована аппаратно с помощью разработанного ПАК. Для этого было выполнено маппирование весовых коэффициентов, которое осуществлялось методом описанном в п. 4.3.1.

После прогона работы аппаратно реализованной ИНСМ доля её правильных исходов (A_H) на тестовой выборке равна 0,767. Если сравнивать A_M, оцениваемую с помощью моделирования с A_H, то можно заметить, что разница между ними составляет 0,001.

Далее была выполнена оценка вклада погрешностей вычислений вызванных ограничением максимально допустимых напряжений, подаваемых на вход кроссбара мемристивных устройств. Для этого выполнялась оценка доли правильных исходов ИНСМ с учетом погрешности веса с помощью моделирования

по алгоритму 3 для 1000 экспериментов, при этом ограничение максимально допустимых напряжений подаваемых на вход кроссбара не учитывалось, то есть $T = \infty$, а $K = 1$. В результате такого исследования было выяснено что доля правильных исходов ИНСМ при такой реализации составит в среднем 0,866.

Основываясь на данных моделирования, было сделано предположение о том, что больший вклад в снижение точности работы вносят именно погрешности, связанные с ограничением максимально допустимых напряжений, подаваемых на вход кроссбара. В связи с этим было решено изменить значение K, для уменьшения влияния данного типа погрешностей на итоговую точность.

Расчет масштабирующего коэффициента K выполнялся аналитически в соответствии с пунктом 2.3. В данном случае $K = 0{,}114$.

Выполнив инференс работы аппаратно реализованной ИНСМ с $K = 0{,}114$, A_H составил 0,667 на тестовой выборке. Как можно заметить из полученных данных, доля правильных исходов ИНСМ снизилась, несмотря на то что был расчет на обратный эффект. Такой результат, предположительно, может быть связан с тем, что из-за масштабирования выходов сверточного слоя, которые затем используются в качестве входов полносвязного слоя, они преобразовались в очень маленькие значения, в результате чего, шумы, имеющиеся в аналоговой электронике, стали оказывать сильное влияние и привели к ещё большему снижению A_H ИНСМ, что подтверждается также данными моделирования.

В связи с этим был выполнен экспериментальный поиск K с применением методов теории планирования эксперимента. Для этого был сформирован план однофакторного эксперимента с различными значениями масштабирующих коэффициентов от 0,2 до 0,6 с шагом 0,1 который представлен в таблице 4.5.

В соответствии с планом эксперимента было выполнено компьютерное моделирование работы ИНСМ для разных значений масштабирующего коэффициента K с применением пороговой функции и рассчитана A_M.

Таблица 4.5 – План эксперимента

Номер эксперимента	**Значение масштабирующего коэффициента**
1	0,2
2	0,3
3	0,4
4	0,5
5	0,6

Результаты моделирования представлены в таблице 4.6. Из полученных результатов видно, что наибольшее значение A_M достигается при масштабирующем коэффициенте 0,5 в 4 эксперименте, при этом A_M в таком случае достигает 0,8.

Таблица 4.6 – Результаты компьютерного моделирования ИНСМ

Номер эксперимента	$\boldsymbol{A_M}$
1	0,773
2	0,799
3	0,799
4	0,800
5	0,769

Соответственно для достижения лучшей A_M нужно использовать $K = 0{,}5$. Для проверки данного предположения был выполнен прогон аппаратно реализованной ИНСМ с данным масштабирующим коэффициентом, в результате чего A_H составила 0,801 %. Таким образом, это наглядно демонстрирует возможности подхода по уменьшению влияния погрешностей связанных с ограничением максимально допустимых напряжений, подаваемых на вход кроссбара мемристивных устройств. Применение данного подхода позволило увеличить A_H 0,767 до 0,801, то есть на 0,34 %, что составляет более 30% от общего снижения A_H.

Также найдена относительная ошибка оценки данной метрики по формуле (14).

$$\delta_M = |A_H - A_M|/A_H \times 100, \tag{4.4}$$

В результате расчетов по феруле (14) было вычислено что относительная ошибка составляет 0,125 %. Полученный результат еще раз показывает качество разработанных моделей и алгоритмов моделирования.

4.3.3 Сводные результаты исследований

По аналогии с описанными в п. 4.3.1 и 4.3.2 были проведены исследования и для других общепринятых тестовых наборов данных, таких как «классификация грибов» (Secondary Mushroom Dataset) и «выявление болезни Паркинсона» (Parkinson Dataset (SST)). Результаты приведены в сводной таблице 4.7.

Таблица 4.7 – Результаты экспериментальных исследований

Задача	**Архитектура ИНСМ**	A_S	A_H	δ_M, %	
				Модель ВАХ	**Алгоритм 3**
Классификация ирисов (Iris Dataset)	Двухслойная сеть прямого распространения	1,000	1,000	0	0
Классификация грибов (Secondary Mushroom Dataset)	Трехслойная сеть прямого распространения	0,939	0,540	24,629	2,407
Классификация элементов гардероба (Fashion MNIST Dataset)	Двухслойная сверточная нейронная сеть	0,867	0,801	6,554	0,125
Выявление болезни Паркинсона (Parkinson Dataset (SST))	Трехслойная рекуррентная нейронная сеть	0,720	0,705	2,178	1,135

4.4 Выводы по главе

1) Выполнена апробация алгоритма моделирования зависимости сопротивления мемристивного устройства от параметров сигналов его задания на примере мемристоров, реализованных на основе тонкопленочной структуры «металл-диэлектрик-металл» в которых, в качестве диэлектрического материала используется стабилизированный иттрием диоксид циркония ZrO2(Y). Построение модели выполнялось на основе данных двухфакторного эксперимента, с применением линейной интерполяции. В результате была получена модель,

которая описывает вариации сопротивлений мемристивного устройства в диапазоне от 9079 до 72225 Ом, а также может показать какие параметры сигнала нужно использовать для достижения требуемых значений сопротивлений.

2) Выполнена апробация алгоритма моделирования зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса. Построение модели выполнялось на основе данных однофакторного эксперимента, полученных из разработанной модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания, с применением линейной интерполяции. В результате была получена модель, которая описывает вариации веса синапса в диапазоне от 0,039 до 0,248.

3) Разработана ИНС в архитектуре многослойного персептрона для решения задачи по классификации на наборе данных «Ирисы Фишера». Для данной ИНС было выполнено компьютерное моделирование с применением разработанных алгоритмов и моделей для оценки функциональной корректности. В итоге прогнозируемая доля правильных исходов составила 1 на тестовых данных. Для проверки полученных данных ИНС была реализована аппаратно на мемристорах. В итоге доля правильных исходов аппаратно реализованной ИНСМ составила 1 на тестовых данных, что полностью соответствует результатам моделирования и показывает эффективность разработанных моделей и алгоритмов.

4) Разработана ИНС в архитектуре сверточной нейронной сети для решения задачи по классификации на наборе данных содержащих изображения элементов гардероба (Fashion MNIST). Для данной ИНС было выполнено компьютерное моделирование с применением разработанных алгоритмов и моделей для оценки функциональной корректности. В итоге прогнозируемая доля правильных исходов составила 0,766 на тестовых данных. Для проверки полученных данных ИНС была реализована аппаратно на мемристорах. В итоге доля правильных исходов составила 0,767 на тестовых данных. Разница в оценке ФК между ними составляет 0,001. Далее была выполнена оценка вклада погрешностей вычислений вызванных ограничением максимально допустимых напряжений, подаваемых на вход кроссбар-массива мемристивных устройств с помощью моделирования. В

результате такого исследования было выяснено, что она примерно составляет более 0,09. Для уменьшения её влияния на итоговую точность с помощью компьютерного моделирования был определен коэффициент масштабирования K в значении 0,5 при котором доля плавильных исходов модели составила 0,8. Прогон аппаратно реализованной ИНСМ с данных коэффициентом масштабирования позволил достигнуть доли правильных исходов в 0,801. Полученный результат экспериментально показывает качество разработанных моделей и алгоритмов, так как разница в точности между моделью и реальным устройством составляет всего 0,001.

5) Подобные эксперименты были проведены для разных общепринятых тестовых задач и разных архитектур нейронных сетей, результаты которых представлены в сводной таблице 4.7. Отличие между результатами моделирования полученными с помощью алгоритма 3 и экспериментом составляет не более 3-х процентов, при этом применение классического подхода к оценки ФК с применением модели ВАХ имеет более высокую ошибку.

Данные результаты получены автором лично и были частично опубликованы в соавторстве (вклад автора более 50 %) в [100,101,111,112].

ЗАКЛЮЧЕНИЕ

1) Разработана модель и алгоритм моделирования зависимости сопротивления мемристивного устройства от параметров сигналов. Показано, что данная модель применима для расчета сопротивления мемристора в зависимости от значений параметров сигнала, и для расчета значений параметров сигнала для заданного значения сопротивления. На основании проведенного вычислительного эксперимента подтверждено, что результаты моделирования совпадают с экспериментальными данными с доверительной вероятностью 95 %. Данные результаты получены автором лично и были частично опубликованы в соавторстве (вклад автора более 50 %) в [100,101,111,112].

2) Разработана модель и алгоритм моделирования зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса. Показано, что данная модель применима для расчета веса синапса нейрона в зависимости от значений сопротивления мемристора, и для расчета значений сопротивления мемристора для заданного значения веса синапса нейрона. На основании проведенного вычислительного эксперимента подтверждено, что результаты моделирования совпадают с экспериментальными данными с доверительной вероятностью 95 %. Данные результаты получены автором лично и были частично опубликованы в соавторстве (вклад автора более 50 %) в [100,101,111,112].

3) Разработан алгоритм оценки ФК ИНСМ с учетом выбранных параметров сигналов задания сопротивлений мемристивных устройств и параметров реально заданных сопротивлений, схемы формирования веса и максимально допустимых напряжений на выходе нейронов. Показано, что применение компьютерного моделирования и предложенного алгоритма позволяют рассчитать значения метрик оценки ФК проектируемой ИНСМ. Данные результаты получены автором лично и были частично опубликованы в соавторстве (вклад автора более 50 %) в [100,101,111,112].

4) Разработан программно-аппаратный комплекс для сбора и накопления экспериментальных данных с мемристивных устройств в соответствии с планами экспериментов и оценки ФК ИНСМ. В данном комплексе реализован функционал автоматизации процесса построения предлагаемых моделей для кроссбар-массивов МУ 32×8 1T1R. Погрешность измерений сопротивления в диапазоне от 500 Ом до 10 кОм составляет не более 1 %. Данные результаты получены автором лично и были частично опубликованы в соавторстве (вклад автора более 50 %) в [100,101,109–115].

5) Проведено сравнение результатов компьютерного моделирования с результатами экспериментов. Для этого аппаратно реализованы 4 разных по архитектуре ИНСМ (полносвязная прямого распространения, сверточная, рекуррентная) с применением кроссбар-массивов 32×8 1T1R на основе диоксида циркония. Показано, что отличие между результатами моделирования и экспериментом составляет не более 3% для предложенного алгоритма и до 25% для существующего метода. Данные результаты получены автором лично и были частично опубликованы в соавторстве (вклад автора более 50 %) в [100,101,111,112].

6) Полученные результаты диссертационного исследования и разработанный ПАК в настоящее время применяются для исследования МУ и проектирования ИНСМ в лаборатории разработки систем искусственного интеллекта МИ ВлГУ, лаборатории мемристорной наноэлектроники ННГУ им. Н.И. Лобачевского и в производственном процессе компании ООО «Поликетон», разрабатывающей нейроморфные системы.

7) В будущем планируется разработка моделей, учитывающих большее число параметров сигнала задания сопротивлений. Данная задача направлена на решение ключевой проблемы мемристивной электроники – воспроизводимости и точности установки весовых коэффициентов, которая напрямую влияет на производительность и надёжность нейроморфных систем. Это позволит ускорить цикл исследований и разработки новых материалов и структур мемристоров в интересах отечественной микроэлектроники.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

В данной диссертационной работе применяются следующие сокращения:

АЦП – аналого-цифровой преобразователь;

БС – белый список;

ВАК – высшая аттестационная комиссия;

ВАХ – вольт-амперная характеристика;

ВлГУ – Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых;

ИИ – искусственный интеллект;

ИНС – искусственная нейронная сеть;

ИНСМ – искусственная нейронная сеть на базе мемристивных устройств;

КМОП - комплементарная структура металл-оксид-полупроводник;

ККИ – кусочно-кубическая интерполяция;

КЛИ – кусочно-линейная интерполяция;

МГППУ – Московский государственный психолого-педагогический университет;

МИ ВлГУ – Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»;

МК – микроконтроллер;

МО – математическое ожидание;

МФТИ – Московский физико-технический институт;

МУ – мемристивное устройство;

НГТУ – Нижегородский государственный технический университет;

НИР – научно-исследовательская работа;

НИЦ – Национальные исследовательский центр;

НОЦ ФТНС ННГУ – научно-образовательный центр «Физика твердотельных наноструктур» Национальный исследовательский Нижегородский государственный университет имени Н. И. Лобачевского;

ООО – общество с ограниченной ответственностью;

ПАК – программно-аппаратный комплекс;

ПО – программное обеспечение;

РФ – Российская Федерация;

РФФИ – Российский фонд фундаментальных исследований;

РНФ – Российский научный фонд;

СВС – состояние высокого сопротивления;

СКО – среднеквадратическое отклонение;

СНС – состояние низкого сопротивления;

СТШ – случайный телеграфный шум;

ЦАП – цифро-аналоговый преобразователь;

ЭВМ – электронная вычислительная машина;

ФК – функциональная корректность;

DRAM – dynamic random-access memory (динамическая память с произвольным доступом);

FLASH – flash memory (флеш-память);

GPU – graphics processing unit (графический процессор);

PCM – phase-change memory (фазоизменяемая память);

ReRAM (RRAM) – resistive random-access memory (резистивная память с произвольным доступом);

SRAM – static random-access memory (статическая память с произвольным доступом);

STT-MRAM – spin-transfer torque magnetic random-access memory (магниторезистивная память с произвольным доступом на основе спинового переноса момента);

TOPS – tera operations per second (триллион операций в секунду).

СПИСОК ЛИТЕРАТУРЫ

1. Burr, G. W. Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element / G. W. Burr [et al.] // IEEE Transactions on Electron Devices. – 2015. – Т. 62, № 11. – С. 3498–3507. – Текст : непосредственный. – DOI: 10.1109/TED.2015.2439635

2. Chai, Z. The Over-Reset Phenomenon in Ta2O5 RRAM Device Investigated by the RTN-Based Defect Probing Technique / Z. Chai [et al.] // IEEE Electron Device Letters. – 2018. – Т. 39, № 7. – С. 955–958. – Текст : непосредственный. – DOI: 10.1109/LED.2018.2833149

3. Roldan, J. B. Variability in Resistive Memories / J. B. Roldan [et al.] // Advanced Intelligent Systems. – 2023. – Т. 5. – С. 2200338. – Текст : непосредственный. – DOI: 10.1002/aisy.202200338

4. Yang, K. Nonlinearity in Memristors for Neuromorphic Dynamic Systems / K. Yang [et al.] // Small Science. – 2021. – Т. 2. – Текст : непосредственный. – DOI: 10.1002/smsc.202100049

5. Mehonic, A. Roadmap to neuromorphic computing with emerging technologies / A. Mehonic [et al.] // APL Materials. – 2024. – Т. 12. – Текст : непосредственный. – DOI: 10.1063/5.0179424

6. Horowitz, M. 1.1 Computing’s energy problem (and what we can do about it) / M. Horowitz // 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC). – IEEE, 2014. – С. 10–14. – Текст : непосредственный. – DOI: 10.1109/ISSCC.2014.6757323

7. Rafiq, S. Investigation of ReRAM Variability on Flow-Based Edge Detection Computing Using HfO2-Based ReRAM Arrays / S. Rafiq [et al.] // IEEE Transactions on Circuits and Systems I: Regular Papers. – 2021. – Т. 68, № 7. – С. 2900–2910. – Текст : непосредственный. – DOI: 10.1109/TCSI.2021.3072210

8. Mehonic, A. Brain-inspired computing needs a master plan / A. Mehonic, A. J. Kenyon // Nature. – 2022. – Т. 604. – С. 255–260. – Текст : непосредственный. – DOI: 10.1038/s41586-021-04362-w

9. Rajput, A. An Energy-Efficient Hybrid SRAM-Based In-Memory Computing Macro for Artificial Intelligence Edge Devices / A. Rajput, A. Tiwari, M. Pattanaik // Circuits, Systems, and Signal Processing. – 2023. – Т. 42. – Текст : непосредственный. – DOI: 10.1007/s00034-022-02284-0

10. Cassidy, A. S. 11.4 IBM NorthPole: An Architecture for Neural Network Inference with a 12nm Chip / A. S. Cassidy [et al.] // 2024 IEEE International Solid-State Circuits Conference (ISSCC). – IEEE, 2024. – Т. 67. – С. 214–215. – Текст : непосредственный. – DOI: 10.1109/ISSCC49657.2024.10454451

11. Davies, M. Loihi: A Neuromorphic Manycore Processor with On-Chip Learning / M. Davies [et al.] // IEEE Micro. – 2018. – Т. 38, № 1. – С. 82–99. – Текст : непосредственный. – DOI: 10.1109/MM.2018.112130359

12. Akopyan, F. TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip / F. Akopyan [et al.] // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. – 2015. – Т. 34, № 10. – С. 1537–1557. – Текст : непосредственный. – DOI: 10.1109/TCAD.2015.2474396

13. Mikhaylov, A. Neuromorphic Computing Based on CMOS-Integrated Memristive Arrays: Current State and Perspectives / A. Mikhaylov [et al.] // Supercomputing Frontiers and Innovations. – 2023. – Т. 10. – С. 77–103. – Текст : непосредственный. – DOI: 10.14529/jsfi230206

14. Pei, J. Towards artificial general intelligence with hybrid Tianjic chip architecture / J. Pei [et al.] // Nature. – 2019. – Т. 572. – С. 106–111. – Текст : непосредственный. – DOI: 10.1038/s41586-019-1424-8

15. Biswas, A. CONV-SRAM: An Energy-Efficient SRAM With In-Memory Dot-Product Computation for Low-Power Convolutional Neural Networks / A. Biswas, A. P. Chandrakasan // IEEE Journal of Solid-State Circuits. – 2019. – Т. 54, № 1. – С. 217–230. – Текст : непосредственный. – DOI: 10.1109/JSSC.2018.2880918

16. Alnatsheh, N. A Novel 8T XNOR-SRAM: Computing-in-Memory Design for Binary/Ternary Deep Neural Networks / N. Alnatsheh [et al.] // Electronics. – 2023. – Т. 12. – С. 877. – Текст : непосредственный. – DOI: 10.3390/electronics12040877

17. Song, J. TD-SRAM: Time-Domain-Based In-Memory Computing Macro for Binary Neural Networks / J. Song [et al.] // IEEE Transactions on Circuits and Systems I: Regular Papers. – 2021. – Т. 68, № 8. – С. 3377–3387. – Текст : непосредственный. – DOI: 10.1109/TCSI.2021.3083275

18. Cho, S. McDRAM v2: In-Dynamic Random Access Memory Systolic Array Accelerator to Address the Large Model Problem in Deep Neural Networks on the Edge / S. Cho [et al.] // IEEE Access. – 2020. – Т. 8. – С. 135223–135243. – Текст : непосредственный. – DOI: 10.1109/ACCESS.2020.3011265

19. Shim, W. Technological Design of 3D NAND-Based Compute-in-Memory Architecture for GB-Scale Deep Neural Network / W. Shim, S. Yu // IEEE Electron Device Letters. – 2021. – Т. 42, № 2. – С. 160–163. – Текст : непосредственный. – DOI: 10.1109/LED.2020.3048101

20. Gallo, M. L. A 64-core mixed-signal in-memory compute chip based on phase-change memory for deep neural network inference / M. L. Gallo [et al.] // Nature Electronics. – 2023. – Т. 6. – С. 680–693. – Текст : непосредственный. – DOI: 10.1038/s41928-023-01010-1

21. Pham, T.-N. STT-BNN: A Novel STT-MRAM In-Memory Computing Macro for Binary Neural Networks / T.-N. Pham [et al.] // IEEE Journal on Emerging and Selected Topics in Circuits and Systems. – 2022. – Т. 12, № 2. – С. 569–579. – Текст : непосредственный. – DOI: 10.1109/JETCAS.2022.3169759

22. Nguyen, V.-T. STT-BSNN: An In-Memory Deep Binary Spiking Neural Network Based on STT-MRAM / V.-T. Nguyen [et al.] // IEEE Access. – 2021. – Т. 9. – С. 151373–151385. – Текст : непосредственный. – DOI: 10.1109/ACCESS.2021.3125685

23. Angizi, S. MRIMA: An MRAM-Based In-Memory Accelerator / S. Angizi [et al.] // IEEE Transactions on Computer-Aided Design of Integrated Circuits and

Systems. – 2020. – Т. 39, № 5. – С. 1123–1136. – Текст : непосредственный. – DOI: 10.1109/TCAD.2019.2907886

24. Rao, M. Thousands of conductance levels in memristors integrated on CMOS / M. Rao [et al.] // Nature. – 2023. – Т. 615. – С. 823–829. – Текст : непосредственный. – DOI: 10.1038/s41586-023-05759-5

25. Wan, W. A compute-in-memory chip based on resistive random-access memory / W. Wan [et al.] // Nature. – 2022. – Т. 608. – С. 504–512. – Текст : непосредственный. – DOI: 10.1038/s41586-022-04992-8

26. Mochida, R. A 4M Synapses integrated Analog ReRAM based 66.5 TOPS/W Neural-Network Processor with Cell Current Controlled Writing and Flexible Network Architecture / R. Mochida [et al.] // 2018 IEEE Symposium on VLSI Technology. – IEEE, 2018. – С. 175–176. – Текст : непосредственный. – DOI: 10.1109/VLSIT.2018.8510676

27. Chen, W.-H. CMOS-integrated memristive non-volatile computing-in-memory for AI edge processors / W.-H. Chen [et al.] // Nature Electronics. – 2019. – Т. 2. – С. 420–428. – Текст : непосредственный. – DOI: 10.1038/s41928-019-0288-0

28. Hung, J.-M. A four-megabit compute-in-memory macro with eight-bit precision based on CMOS and resistive random-access memory for AI edge devices / J.-M. Hung [et al.] // Nature Electronics. – 2021. – Т. 4. – С. 921–930. – Текст : непосредственный. – DOI: 10.1038/s41928-021-00676-9

29. Cai, F. A Fully Integrated System-on-Chip Design with Scalable Resistive Random-Access Memory Tile Design for Analog in-Memory Computing / F. Cai [et al.] // Advanced Intelligent Systems. – 2022. – Т. 4, № 8. – С. 2200063. – Текст : непосредственный. – DOI: 10.1002/aisy.202200014

30. Yao, P. Fully hardware-implemented memristor convolutional neural network / P. Yao [et al.] // Nature. – 2020. – Т. 577. – С. 641–646. – Текст : непосредственный. – DOI: 10.1038/s41586-020-1942-4

31. Xu, J. A Cascaded ReRAM-based Crossbar Architecture for Transformer Neural Network Acceleration / J. Xu [et al.] // ACM Transactions on Design Automation of Electronic Systems. – 2024. – Текст : непосредственный. – DOI: 10.1145/3701034

32. Zheng, Y.-L. An Energy-Efficient Inference Engine for a Configurable ReRAM-Based Neural Network Accelerator / Y.-L. Zheng [et al.] // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. – 2022. – Т. 41, № 8. – С. 2689–2700. – Текст : непосредственный. – DOI: 10.1109/TCAD.2022.3184464

33. Ogbogu, C. Data Pruning-enabled High Performance and Reliable Graph Neural Network Training on ReRAM-based Processing-in-Memory Accelerators / C. Ogbogu [et al.] // ACM Transactions on Design Automation of Electronic Systems. – 2024. – Т. 29, № 3. – С. 1–23. – Текст : непосредственный. – DOI: 10.1145/3656171

34. Liu, B. Frequency-Domain Inference Acceleration for Convolutional Neural Networks Using ReRAMs / B. Liu [et al.] // IEEE Transactions on Parallel and Distributed Systems. – 2023. – Т. 34, № 6. – С. 1893–1906. – Текст : непосредственный. – DOI: 10.1109/TPDS.2023.3322907

35. Azamat, A. Partial Sum Quantization for Reducing ADC Size in ReRAM-based Neural Network Accelerators / A. Azamat [et al.] // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. – 2023. – Т. 42, № 8. – С. 2693–2706. – Текст : непосредственный. – DOI: 10.1109/TCAD.2023.3294461

36. Yousuf, O. Device Modeling Bias in ReRAM-based Neural Network Simulations / O. Yousuf [et al.] // IEEE Journal on Emerging and Selected Topics in Circuits and Systems. – 2023. – Т. 13, № 2. – С. 477–488. – Текст : непосредственный. – DOI: 10.1109/JETCAS.2023.3238295

37. Stecconi, T. Filamentary TaOx/HfO2 ReRAM Devices for Neural Networks Training with Analog In-Memory Computing / T. Stecconi [et al.] // Advanced Electronic Materials. – 2022. – Т. 8, № 6. – С. 2101037. – Текст : непосредственный. – DOI: 10.1002/aelm.202200448

38. Wang, W. Trained Biased Number Representation for ReRAM-Based Neural Network Accelerators / W. Wang, B. Lin // ACM Journal on Emerging Technologies in Computing Systems. – 2019. – Т. 15, № 3. – С. 1–17. – Текст : непосредственный. – DOI: 10.1145/3304107

39. Ji, Y. A Reduced Architecture for ReRAM-Based Neural Network Accelerator and Its Software Stack / Y. Ji, Z. Liu, Y. Zhang // IEEE Transactions on

Computers. – 2020. – Т. 69, № 8. – С. 1173–1186. – Текст : непосредственный. – DOI: 10.1109/TC.2020.2988248

40. Lin, J. Analysis and Simulation of Capacitor-Less ReRAM-Based Stochastic Neurons for the in-Memory Spiking Neural Network / J. Lin, J.-S. Yuan // IEEE Transactions on Biomedical Circuits and Systems. – 2018. – Т. 12, № 5. – С. 1104–1117. – Текст : непосредственный. – DOI: 10.1109/TBCAS.2018.2843286

41. Kim, H. ADC-Free ReRAM-Based In-Situ Accelerator for Energy-Efficient Binary Neural Networks / H. Kim, Y. Jung, L.-S. Kim // IEEE Transactions on Computers. – 2022. – Т. 71, № 3. – С. 620–632. – Текст : непосредственный. – DOI: 10.1109/TC.2022.3224800

42. Gi, S. A ReRAM-Based Convolutional Neural Network Accelerator Using the Analog Layer Normalization Technique / S. Gi [et al.] // IEEE Transactions on Industrial Electronics. – 2022. – Т. 69, № 12. – С. 13378–13387. – Текст : непосредственный. – DOI: 10.1109/TIE.2022.3190876

43. Wang, W. Computing of Temporal Information in Spiking Neural Networks with ReRAM Synapses / W. Wang [et al.] // Faraday Discussions. – 2018. – Т. 213. – С. 483–497. – Текст : непосредственный. – DOI: 10.1039/C8FD00097B

44. Lin, J. CNNWire: Boosting Convolutional Neural Network with Winograd on ReRAM based Accelerators / J. Lin [et al.] // 2019 56th ACM/IEEE Design Automation Conference (DAC). – IEEE, 2019. – С. 1–6. – Текст : непосредственный. – DOI: 10.1145/3299874.3318018

45. Chen, Y. A ReRAM-Based Row-Column-Oriented Memory Architecture for Convolutional Neural Networks / Y. Chen [et al.] // IEICE Transactions on Electronics. – 2019. – Т. E102.C, № 7. – С. 580–584. – Текст : непосредственный. – DOI: 10.1587/transele.2018CTS0001

46. Amirsoleimani, A. In-Memory Vector-Matrix Multiplication in Monolithic Complementary Metal–Oxide–Semiconductor-Memristor Integrated Circuits: Design Choices, Challenges, and Perspectives / A. Amirsoleimani [et al.] // Advanced Intelligent Systems. – 2020. – Т. 2, № 10. – С. 2000115. – Текст : непосредственный. – DOI: 10.1002/aisy.202000115

47. Lapkin, D. An Organic Memristive Element Based on Single Polyaniline/Polyamide-6 Fiber / D. Lapkin [et al.] // Technical Physics Letters. – 2017. – Т. 43, № 12. – С. 1102–1104. – Текст : непосредственный. – DOI: 10.1134/S1063785017120227

48. Malakhova, Y. Planar and 3D fibrous polyaniline-based materials for memristive elements / Y. Malakhova [et al.] // Soft Matter. – 2017. – Т. 13, № 43. – С. 7888–7896. – Текст : непосредственный. – DOI: 10.1039/c7sm01773a

49. Zhou, Y. Associative Memory for Image Recovery with a High-Performance Memristor Array / Y. Zhou [et al.] // Advanced Functional Materials. – 2019. – Т. 29, № 29. – С. 1900158. – Текст : непосредственный. – DOI: 10.1002/adfm.201900155

50. Brzhezinskaya, M. Large-scalable graphene oxide films with resistive switching for non-volatile memory applications / M. Brzhezinskaya [et al.] // Journal of Alloys and Compounds. – 2020. – Т. 849. – С. 156699. – Текст : непосредственный. – DOI: 10.1016/j.jallcom.2020.156699

51. Смирнов, В. А. Исследование мемристорного эффекта в нанокристаллических пленках ZnO / В. А. Смирнов [и др.] // Журнал технической физики. – 2019. – Т. 89, № 1. – С. 77–82. – Текст : непосредственный.

52. Andreeva, N. V. Multilevel resistive switching in TiO 2 /Al 2 O 3 bilayers at low temperature / N. V. Andreeva, A. Ivanov, A. Petrov // AIP Advances. – 2018. – Т. 8, № 2. – С. 025208. – Текст : непосредственный. – DOI: 10.1063/1.5019570

53. Levanov, V. Memristive Properties of Structures Based on (Co41Fe39B20)x(LiNbO3)100–x Nanocomposites / V. Levanov [et al.] // Journal of Communications Technology and Electronics. – 2018. – Т. 63, № 5. – С. 491–496. – Текст : непосредственный. – DOI: 10.1134/S1064226918050078

54. Li, C. Three-dimensional crossbar arrays of self-rectifying Si/SiO2/Si memristors / C. Li [et al.] // Nature Communications. – 2017. – Т. 8. – С. 15666. – Текст : непосредственный. – DOI: 10.1038/ncomms15666

55. Степанов, А. В. Улучшение Параметров Мемристоров На Основе Оксида Кремния Методом Ионного Облучения / А. В. Степанов [и др.] // Вестник

Чувашской Государственной Сельскохозяйственной Академии. – 2018. – № 1 (4). – С. 87–91. . – Текст : непосредственный.

56. Mikheev, V. Ferroelectric Second-Order Memristor / V. Mikheev [et al.] // ACS Applied Materials & Interfaces. – 2019. – Т. 11, № 35. – С. 32108–32114. – Текст : непосредственный. – DOI: 10.1021/acsami.9b08189

57. Prezioso, M. Training and Operation of an Integrated Neuromorphic Network Based on Metal-Oxide Memristors / M. Prezioso [et al.] // Nature. – 2014. – Т. 521. – С. 61–64. – Текст : непосредственный. – DOI: 10.1038/nature14441

58. Коряжкина, М. Н. Влияние облучения на параметры резистивного переключения в мемристивных структурах на основе стабилизированного диоксида циркония / М. Н. Коряжкина [и др.] // Всероссийская конференция «Актуальные проблемы твердотельной электроники и микроэлектроники». – Национальный исследовательский ядерный университет «МИФИ», 2019. – С. 202–205. – Текст : непосредственный.

59. Тихов, С. В. Механизмы токопереноса и резистивного переключения в конденсаторах со слоями стабилизированного иттрием диоксида гафния / С. В. Тихов [и др.] // Журнал Технической Физики. – 2019. – Т. 89, № 6. – С. 927–934. – Текст : непосредственный.

60. Герасимова, С. А. Имитация синаптической связи нейроноподобных генераторов с помощью мемристивного устройства (14) / С. А. Герасимова [и др.] // Журнал Технической Физики. – 2017. – Т. 87, № 8. – С. 1248–1254. – Текст : непосредственный.

61. Горшков, О. Н. Резистивное переключение в структурах «металл-диэлектрик-металл» на основе оксида германия и стабилизированного диоксида циркония / О. Н. Горшков [и др.] // Письма В Журнал Технической Физики. – 2014. – Т. 40, № 3. – С. 12–19. – Текст : непосредственный.

62. Joksas, D. Nonideality-Aware Training for Accurate and Robust Low-Power Memristive Neural Networks / D. Joksas [et al.] // Advanced Science. – 2022. – Т. 9, № 36. – С. 2203432. – Текст : непосредственный. – DOI: 10.1002/advs.202105784

63. Merrikh-Bayat, F. Implementation of Multilayer Perceptron Network with Highly Uniform Passive Memristive Crossbar Circuits / F. Merrikh-Bayat [et al.] // Nature Communications. – 2018. – Т. 9. – С. 2331. – Текст : непосредственный. – DOI: 10.1038/s41467-018-04482-4

64. Adhikari, S. P. Building Cellular Neural Network Templates with a Hardware Friendly Learning Algorithm / S. P. Adhikari [et al.] // Neurocomputing. – 2018. – Т. 312. – С. 276–284. – Текст : непосредственный. – DOI: 10.1016/j.neucom.2018.05.113

65. Guo, Y. Unsupervised Learning on Resistive Memory Array Based Spiking Neural Networks / Y. Guo [et al.] // Frontiers in Neuroscience. – 2019. – Т. 13. – С. 812. – Текст : непосредственный. – DOI: 10.3389/fnins.2019.00812

66. Joksas, D. Memristive, Spintronic, and 2D-Materials-Based Devices to Improve and Complement Computing Hardware / D. Joksas [et al.] // Advanced Intelligent Systems. – 2022. – Т. 4, № 5. – С. 2100183. – Текст : непосредственный. – DOI: 10.1002/aisy.202200068

67. Yang, C. A Circuit-Based Neural Network with Hybrid Learning of Backpropagation and Random Weight Change Algorithms / C. Yang [et al.] // Sensors. – 2016. – Т. 17, № 1. – С. 16. – Текст : непосредственный. – DOI: 10.3390/s17010016

68. Joksas, D. Committee machines—a universal method to deal with non-idealities in memristor-based neural networks / D. Joksas [et al.] // Nature Communications. – 2020. – Т. 11. – С. 4273. – Текст : непосредственный. – DOI: 10.1038/s41467-020-18098-0

69. Yakopcic, C. Tolerance to defective memristors in a neuromorphic learning circuit / C. Yakopcic, R. Hasan, T. Taha // 2014 International Joint Conference on Neural Networks (IJCNN). – IEEE, 2014. – С. 249–256. – Текст : непосредственный. – DOI: 10.1109/NAECON.2014.7045810

70. Danilin, S. Design of Multilayer Perceptron Network Based on Metal-Oxide Memristive Devices / S. Danilin [et al.] // 2019 12th International Conference on Developments in eSystems Engineering (DeSE). – IEEE, 2019. – С. 533–538. – Текст : непосредственный. – DOI: 10.1109/DeSE.2019.00103

71. Hasan, R. On-chip training of memristor crossbar based multi-layer neural networks / R. Hasan, T. Taha, C. Yakopcic // Microelectronics Journal. – 2017. – Т. 66. – С. 31–40. – Текст : непосредственный. – DOI: 10.1016/j.mejo.2017.05.005

72. Adhikari, S. P. Memristor Bridge Synapse-Based Neural Network and Its Learning / S. P. Adhikari [et al.] // IEEE Transactions on Neural Networks and Learning Systems. – 2012. – Т. 23, № 9. – С. 1426–1435. – Текст : непосредственный. – DOI: 10.1109/TNNLS.2012.2204770

73. Kim, Y. Shared memristance restoring circuit for memristor-based Cellular Neural Networks / Y. Kim, S. Shin, K.-S. Min // 2014 14th International Workshop on Cellular Nanoscale Networks and their Applications (CNNA). – IEEE, 2014. – С. 1–2. – Текст : непосредственный. – DOI: 0.1109/CNNA.2014.6888625

74. Wang, Z. Reinforcement learning with analogue memristor arrays / Z. Wang [et al.] // Nature Electronics. – 2019. – Т. 2. – С. 115–124. – Текст : непосредственный. – DOI: 10.1038/s41928-019-0221-6

75. Lammie, C. MemTorch: An Open-source Simulation Framework for Memristive Deep Learning Systems / C. Lammie [et al.] // Neurocomputing. – 2022. – Т. 485. – С. 124–133. – Текст : непосредственный. – DOI: 10.1016/j.neucom.2022.02.043

76. Lu, A. NeuroSim Simulator for Compute-in-Memory Hardware Accelerator: Validation and Benchmark / A. Lu [et al.] // Frontiers in Artificial Intelligence. – 2021. – Т. 4. – С. 659060. – Текст : непосредственный. – DOI: 10.3389/frai.2021.659060

77. Busygin, A. Mathematical Model of Metal–Oxide Memristor Resistive Switching based on Full Physical Model of Heat and Mass Transfer of Oxygen Vacancies and Ions / A. Busygin [et al.] // physica status solidi (a). – 2022. – Т. 219, № 21. – С. 2200300. – Текст : непосредственный. – DOI: 10.1002/pssa.202200478

78. Hossen, I. Data-driven RRAM device models using Kriging interpolation / I. Hossen [et al.] // Scientific Reports. – 2022. – Т. 12. – С. 7490. – Текст : непосредственный. – DOI: 10.1038/s41598-022-09556-4

79. Lee, Y. A Compact Memristor Model Based on Physics-Informed Neural Networks / Y. Lee, K. Kim, J. Lee // Micromachines. – 2024. – Т. 15, № 2. – С. 253. – Текст : непосредственный. – DOI: 10.3390/mi15020253

80. Patni, T. VVTEAM: A Compact Behavioral Model for Volatile Memristors / T. Patni, R. Daniels, S. Kvatinsky // arXiv. – 2024. – arXiv:2409.17723. – Текст : непосредственный. – DOI: 10.48550/arXiv.2409.17723

81. Wang, T. A Faithful and Compact Diffusive Memristor Model / T. Wang [et al.] // IEEE Transactions on Circuits and Systems for Artificial Intelligence. – 2024. – Текст : непосредственный. – DOI: 10.1109/TCASAI.2024.3484370

82. Ebrahim, A. Compact multifilament model of resistive switching metal-oxide memristor / A. Ebrahim [et al.] // Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy. – 2023. – Т. 9, № 2. – С. 128–138. – Текст : непосредственный. – DOI: 10.21684/2411-7978-2023-9-2-128-138

83. Zhuo, Y. A Dynamical Compact Model of Diffusive and Drift Memristors for Neuromorphic Computing / Y. Zhuo [et al.] // Advanced Electronic Materials. – 2021. – Т. 8, № 2. – С. 2100817. – Текст : непосредственный. – DOI: 10.1002/aelm.202100696

84. Kim, S. Compact Two-State-Variable Second-Order Memristor Model / S. Kim, H.-D. Kim, S.-J. Choi // Small. – 2016. – Т. 12, № 14. – С. 1890–1899. – Текст : непосредственный. – DOI: 10.1002/smll.201600088

85. Acal, C. Holistic Variability Analysis in Resistive Switching Memories Using a Two-Dimensional Variability Coefficient / C. Acal [et al.] // ACS Applied Materials & Interfaces. – 2023. – Т. 15, № 32. – С. 38759–38771. – Текст : непосредственный. – DOI: 0.1021/acsami.2c22617

86. Malik, A. An Absorbing Markov Chain Model for Stochastic Memristive Devices / A. Malik, C. Papavassiliou, S. Stathopoulos // 2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS). – IEEE, 2022. – С. 1–4. – Текст : непосредственный. – DOI: 10.1109/MOCAST54814.2022.9837672

87. Gambuzza, L. A data driven model of TiO2 printed memristors / L. Gambuzza [et al.] // 2013 8th International Conference on Electrical and Electronics Engineering (ELECO). – IEEE, 2013. – С. 261–264. – Текст : непосредственный. – DOI: 10.1109/ELECO.2013.6713923

88. Saha, S. Experimental Demonstration of SnO_2 Nanofiber-Based Memristors and Their Data-Driven Modeling for Nanoelectronic Applications / S. Saha [et al.] // Chip. – 2023. – Т. 2, № 4. – С. 100075. – Текст : непосредственный. – DOI: 10.1016/j.chip.2023.100075

89. Messaris, Y. A Data-Driven Verilog-A ReRAM Model / Y. Messaris [et al.] // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. – 2018. – Т. 37, № 11. – С. 2681–2692. – Текст : непосредственный. – DOI: 10.1109/TCAD.2018.2791468

90. Lee, J.-H. Exploring Cycle-to-Cycle and Device-to-Device Variation Tolerance in MLC Storage-Based Neural Network Training / J.-H. Lee [et al.] // IEEE Transactions on Electron Devices. – 2019. – Т. 66, № 7. – С. 3089–3095. – Текст : непосредственный. – DOI: 10.1109/TED.2019.2906249

91. Luo, W.-C. Statistical Model and Rapid Prediction of RRAM SET Speed–Disturb Dilemma / W.-C. Luo [et al.] // IEEE Transactions on Electron Devices. – 2013. – Т. 60, № 11. – С. 3760–3766. – Текст : непосредственный. – DOI: 10.1109/TED.2013.2281991

92. Alonso, F. J. Memristor variability and stochastic physical properties modeling from a multivariate time series approach / F. J. Alonso [et al.] // Chaos, Solitons & Fractals. – 2021. – Т. 143. – С. 110461. – Текст : непосредственный. – DOI: 10.1016/j.chaos.2020.110461

93. Imani, M. RAPIDNN: In-Memory Deep Neural Network Acceleration Framework / M. Imani [et al.] // 2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC). – IEEE, 2018. – С. 629–634. – Текст : непосредственный. – DOI: 10.48550/arXiv.1806.05794

94. Ankit, A. PUMA: A Programmable Ultra-efficient Memristor-based Accelerator for Machine Learning Inference / A. Ankit [et al.] // 2019 56th ACM/IEEE Design Automation Conference (DAC). – IEEE, 2019. – С. 1–6. – Текст : непосредственный. – DOI: 10.1145/3297858.3304049

95. Huang, J. A tool for emulating neuromorphic architectures with memristive models and devices / J. Huang [et al.] // 2022 IEEE International Symposium on Circuits

and Systems (ISCAS). – IEEE, 2022. – С. 1092–1096. – Текст : непосредственный. – DOI: 10.1109/ISCAS48785.2022.9937599

96. Xia, L. MNSIM: Simulation Platform for Memristor-Based Neuromorphic Computing System / L. Xia [et al.] // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. – 2018. – Т. 37, № 5. – С. 1009–1022. – Текст : непосредственный. – DOI: 10.3850/9783981537079_0549

97. Peng, X. DNN+NeuroSim V2.0: An End-to-End Benchmarking Framework for Compute-in-Memory Accelerators for On-Chip Training / X. Peng [et al.] // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. – 2021. – Т. 40, № 11. – С. 2306–2319. – Текст : непосредственный. – DOI: 10.1109/tcad.2020.3043731

98. Rasch, M. J. A Flexible and Fast PyTorch Toolkit for Simulating Training and Inference on Analog Crossbar Arrays / M. J. Rasch [et al.] // 2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS). – IEEE, 2021. – С. 1–4. – Текст : непосредственный. – DOI: 10.1109/AICAS51828.2021.9458494

99. Lin, M.-Y. DL-RSIM: A Simulation Framework to Enable Reliable ReRAM-based Accelerators for Deep Learning / M.-Y. Lin [et al.] // 2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD). – IEEE, 2018. – С. 1–8. – Текст : непосредственный. – DOI: 10.1145/3240765.3240800

100. Борданов, И. А. Оценка точности работы искусственных нейронных сетей на базе мемристивных устройств на основе теории планирования эксперимента / И. А. Борданов, Л. Я. Королёв, С. А. Щаников, А. Н. Михайлов // Радиотехнические и телекоммуникационные системы. – 2025. – № 2. – С. 40–52. – Текст : непосредственный.

101. Борданов, И. А. Оценка точности работы искусственных нейронных сетей на базе мемристоров с применением моделей на основе данных / И. А. Борданов, С. А. Щаников // Радиотехнические и телекоммуникационные системы. – 2024. – № 2 (54). – С. 59–68. – Текст : непосредственный.

102. Данилин, С. Н. Количественное определение отказоустойчивости искусственных нейронных сетей на базе мемристоров / С. Н. Данилин, С. А. Щаников, И. А. Борданов, А. Д. Зуев // Нейрокомпьютеры: разработка, применение. – 2020. – Т. 22, № 1. – С. 55–65. – Текст : непосредственный.

103. Борданов, И. А. Современное состояние в области аппаратной реализации искусственных нейронных сетей на базе мемристоров / И. А. Борданов, С. А. Щаников, С. Н. Данилин // Телекоммуникации. – 2020. – № 8. – С. 35–48. – Текст : непосредственный.

104. Mehonic, A. Simulation of Inference Accuracy Using Realistic RRAM Devices / A. Mehonic [et al.] // Frontiers in Neuroscience. – 2019. – Т. 13. – Текст : непосредственный. – DOI: 10.3389/fnins.2019.00593

105. The Joglekar Resistance Switch Memristor Model in LTSpice / Knowm. – URL: https://knowm.org/the-joglekar-resistance-switch-memristor-model-in-ltspice/ (дата обращения: 09.11.2024). – Текст : электронный.

106. Bordanov, I. A. Determining the fault tolerance of memristorsbased neural network using simulation and design of experiments / I. A. Bordanov [et al.] // 2018 Engineering and telecommunication (EnT-MIPT). – IEEE, 2018. – P. 205-209. – Текст : непосредственный. – DOI: 10.1109/EnT-MIPT.2018.00053.

107. Bordanov, I. A. High-performance software for memristor-based neural network simulation and optimization / I. A. Bordanov, R. A. Mineev, S. N. Danilin. – Текст : непосредственный // 2021 International Conference Engineering and Telecommunication (En&T) / Moscow Institute of Physics and Technology – IEEE, 2021. – P. 1–4.

108. Bordanov, I. A. Modeling and hardware implementation of vector-matrix multiplier based on 32x8 1T1R memristive crossbar array / I. A. Bordanov [et al.] // 2023 7th Scientific School Dynamics of Complex Networks and their Applications (DCNA). – IEEE, 2023. – P. 249–251. – Текст : непосредственный. – DOI: 10.1109/DCNA59899.2023.10290511.

109. Bordanov, I. A. Simulation of calculation errors in memristive crossbars for artificial neural networks / I. A. Bordanov, A. A. Antonov, L. Ya. Korolev // 2023

International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). – IEEE, 2023. – P. 1008–1012. – Текст : непосредственный. – DOI: 10.1109/ICIEAM57311.2023.10139308.

110. Борданов, И. А. Исследование влияния погрешностей матрично-векторного умножения на точность работы искусственных нейронных сетей на базе мемристоров / И. А. Борданов, С. Н. Данилин – Текст : непосредственный // Нейрокомпьютеры и их применение : сборник тезисов XXI Всероссийской научной конференции / Московский государственный психолого-педагогический университет. – Москва, 2023. – С. 156–157.

111. Борданов, И. А. Применение методологии имитационного моделирования для оценки точности работы искусственных нейронных сетей на базе мемристивных устройств / И. А. Борданов, С. А. Щаников – Текст : непосредственный // Труды первой школы-конференции с международным участием «Нейроэлектроника и нейротехнологии будущего» / ННГУ. – Нижний Новгород, 2024. – С. 32.

112. Борданов, И. А. Оценка точности работы искусственных нейронных сетей на базе мемристоров с применением моделей на основе данных / И. А. Борданов, С. А. Щаников. – Текст : непосредственный // Информационные системы и технологии - 2025 : программа и аннотации докладов XXXI Международной научно-технической конференции / НГТУ им. Р. Е. Алексеева. – Нижний Новгород, 2025. – С. 36.

113. Свидетельство № 2023666086 Российская Федерация. Программа для моделирования матрично-векторного умножения с учетом погрешностей мемристивных устройств : № 2023665065 : заявлено 17.07.2023 : опубликовано 26.07.2023 / Борданов И. А., Щаников С. А. – 1 с. – Текст : непосредственный.

114. Свидетельство № 2024619751 Российская Федерация. Программа для оценки точности работы искусственных нейронных сетей на базе мемристоров с учетом погрешности матрично-векторного умножения : № 2024616902 : заявлено 02.04.2024 : опубликовано 25.04.2024 / Борданов И. А., Щаников С. А. – 1 с. – Текст : непосредственный.

115. Свидетельство № 2019661246 Российская Федерация. Модуль определения точности функционирования искусственных нейронных сетей на базе мемристоров для системы имитационного моделирования : № 2019619978 : заявлено 12.08.2019 : опубликовано 23.08.2019 / Щаников С. А., Борданов И. А., Данилин С. А., Зуев А. Д. – 1 с. – Текст : непосредственный.

Приложение А

(обязательное)

Акты внедрения

«УТВЕРЖДАЮ»

Директор МИ ВлГУ, д.т.н., профессор

_____ А.Л. Жизняков

«01» _____ 2026 г.

АКТ

об использовании результатов диссертационной работы «Модели и алгоритмы оценки функциональной корректности искусственных нейронных сетей на базе мемристоров»
Борданова Ильи Алексеевича

Мы, нижеподписавшиеся, декан факультета информационных технологий и радиоэлектроники, к.т.н. Рыжкова Мария Николаевна, заведующий кафедрой информационных систем, д.т.н. Андрианов Дмитрий Евгеньевич, составили настоящий акт о том, что результаты диссертационной работы Борданова И.А. внедрены в учебный процесс кафедры информационных систем и в рабочий процесс лаборатории систем искусственного интеллекта МИ ВлГУ. Используются разделы, связанные с алгоритмами и моделями оценки функциональной корректности искусственных нейронных сетей на базе мемристивных устройств (МУ), а также программно-аппаратный комплекс для исследования характеристик МУ в активных кроссбар-массивах в архитектуре 1T1R.

Декан ФИТР, к.т.н. М.Н. Рыжкова

Зав. кафедрой ИС, д.т.н. Д.Е. Андрианов

Рисунок А.1 – Акт внедрения из лаборатории разработки систем искусственного интеллекта МИ ВлГУ

СПРАВКА

о внедрении результатов диссертационного исследования
Борданова Ильи Алексеевича
«Модели и алгоритмы оценки функциональной корректности искусственных нейронных сетей на базе мемристоров»
в ННГУ им. Н.И. Лобачевского

Настоящим Актом удостоверяется, что результаты диссертационной работы Борданова И.А. «Модели и алгоритмы оценки функциональной корректности искусственных нейронных сетей на базе мемристоров» на соискание ученой степени кандидата технических наук в настоящее время используются при выполнении научно-исследовательских работ в ННГУ им. Н.И. Лобачевского.

Предложенные Бордановым И. А. алгоритмы и модели оценки функциональной корректности искусственных нейронных сетей на базе мемристоров использованы в ходе выполнения Проекта 9.1 «Нейроэлектроника – интеллектуальные нейроморфные и нейрогибридные системы на основе новой электронной компонентной базы» Национального центра физики и математики, а разработанный программно-аппаратный комплекс используется в работе НИЛ «Лаборатория мемристорной наноэлектроники» для исследования характеристик и демонстрации работы аппаратно реализованных нейронных сетей на базе кроссбар-массивов мемристивных устройств 32x8 1T1R.

Проректор по науке и инновациям,
ННГУ им. Н.И. Лобачевского М.Ю. Грязнов

М.П.

Директор НОЦ ФТНС,
Руководитель НИР А.Н. Михайлов

Рисунок А.2 – Акт внедрения из лаборатории мемристорной наноэлектроники НОЦ ФТНС ННГУ

Общество с ограниченной ответственностью
«ПОЛИКЕТОН»

603009, г. Нижний Новгород, ул. Батумская, д.7А, пом. 308
ОГРН 1157746483553 ИНН 7723393081 КПП 526101001
www.polyketon.ru e-mail: info@polyketon.ru

АКТ

о внедрении (практическом применении) результатов диссертационного исследования Борданова Ильи Алексеевича на тему «Модели и алгоритмы оценки функциональной корректности искусственных нейронных сетей на базе мемристоров»

Результаты диссертационной работы Борданова И.А. «Модели и алгоритмы оценки функциональной корректности искусственных нейронных сетей на базе мемристоров» на соискание учёной степени кандидата технических наук представляют практический интерес для ООО «ПОЛИКЕТОН».

Разработанный в рамках диссертационного исследования программно-аппаратный комплекс использован для задач входного контроля работоспособности интегральных микросхем, содержащих кроссбар-массивы с мемристивными устройствами в архитектуре 32x8 1T1R.

Генеральный директор
ООО «ПОЛИКЕТОН» А.Ю. Слиняков

«20» января 2026 г

Рисунок А.3 – Акт внедрения из компании ООО «Поликетон»

Приложение Б

(обязательное)

Свидетельства ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2024619751

Программа для оценки точности работы искусственных нейронных сетей на базе мемристоров с учетом погрешности матрично-векторного умножения

Правообладатели: ***Борданов Илья Алексеевич (RU), Щаников Сергей Андреевич (RU)***

Авторы: ***Борданов Илья Алексеевич (RU), Щаников Сергей Андреевич (RU)***

Заявка № **2024616902**

Дата поступления **02 апреля 2024 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***25 апреля 2024 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Ю.С. Зубов

Рисунок Б.1 – Программа для оценки точности работы искусственных нейронных сетей на базе мемристоров с учетом погрешности матрично-векторного умножения

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2023666086

Программа для моделирования матрично-векторного умножения с учетом погрешностей мемристивных устройств

Правообладатели: ***Борданов Илья Алексеевич (RU), Щаников Сергей Андреевич (RU)***

Авторы: ***Борданов Илья Алексеевич (RU), Щаников Сергей Андреевич (RU)***

Заявка № **2023665065**
Дата поступления **17 июля 2023 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***26 июля 2023 г.***

Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

Рисунок Б.2 – Программа для моделирования матрично-векторного умножения с учетом погрешностей мемристивных устройств

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019661246

Модуль определения точности функционирования искусственных нейронных сетей на базе мемристоров для системы имитационного моделирования

Правообладатель: ***Федеральное государственное бюджетное образовательное учреждение высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых» (RU)***

Авторы: ***Данилин Сергей Николаевич (RU), Щаников Сергей Андреевич (RU), Зуев Антон Дмитриевич (RU), Борданов Илья Алексеевич (RU)***

Заявка № **2019619978**

Дата поступления **12 августа 2019 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***23 августа 2019 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

Рисунок Б.3 – Модуль определения точности функционирования искусственных нейронных сетей на базе мемристоров для системы имитационного моделирования