

Нейронные сети и интеллектуальные системы

DOI 10.24412/2221-2574-2025-2-40-52

УДК 004.383.8.032.26

ОЦЕНКА ТОЧНОСТИ РАБОТЫ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ НА БАЗЕ МЕМРИСТИВНЫХ УСТРОЙСТВ НА ОСНОВЕ ТЕОРИИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

Борданов Илья Алексеевич

инженер-исследователь лаборатории разработки систем искусственного интеллекта, Муромский институт¹ (филиал) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: bordanov2011@yandex.ru

Королёв Леонид Ярославович

лаборант-исследователь научно-исследовательской лаборатории стохастических мультстабильных систем отдела фундаментальных и прикладных исследований, ФГАОУ ВО «Национальный исследовательский Нижегородский государственный университет имени Н.И. Лобачевского (ННГУ)².

E-mail: madimtor@yandex.ru

Щаников Сергей Андреевич

кандидат технических наук, доцент, научный руководитель лаборатории разработки систем искусственного интеллекта, Муромский институт¹ (филиал) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: seach@inbox.ru

Михайлов Алексей Николаевич

кандидат физико-математических наук, директор научно-образовательного центра «Физика твердотельных наноструктур» отдела фундаментальных и прикладных исследований, ФГАОУ ВО «Национальный исследовательский Нижегородский государственный университет имени Н.И. Лобачевского (ННГУ)².

E-mail: mian@nifti.unn.ru

¹Адрес: 602264, Российская Федерация, Владимирская обл., г. Муром, Орловская ул., д. 23.

²Адрес: 603022, Российская Федерация, г. Нижний Новгород, просп. Гагарина, д. 23.

Аннотация: На сегодняшний день компьютерное моделирование является одним из наиболее распространённых подходов к оценке точности работы искусственных нейронных сетей на базе мемристивных устройств (ИНСМ) с учётом вариаций их сопротивлений. В процессе такого моделирования обычно используются модели, описывающие связь между приложенным напряжением и током в цепи с мемристором. Эти модели хорошо воспроизводят поведение мемристора, однако при оценке точности ИНСМ имеют ряд недостатков, связанных со сложностью и неоднозначностью в подходах к учёту и заданию вариаций сопротивлений от цикла к циклу и от устройства к устройству, ресурсоёмкостью процесса моделирования ИНСМ и т.д. В данной работе предложен новый подход к оценке точности ИНСМ на этапе исследовательского проектирования, который заключается в том, что точность конкретного варианта ИНСМ нужно оценивать в привязке к параметрам сигнала задания сопротивления мемристивного устройства, на основании установленной взаимосвязи между этими параметрами и реально получаемыми в эксперименте значениями сопротивлений и весов. Данный подход предполагает использование теории планирования эксперимента для создания модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания и модели зависимости весов синапсов ИНСМ от сопротивлений и схемы формирования веса и их применение в компьютерном моделировании при оценке точности работы сети.

Ключевые слова: искусственные нейронные сети, мемристивные устройства, компьютерное моделирование, погрешность вычисления, модели на основе данных.

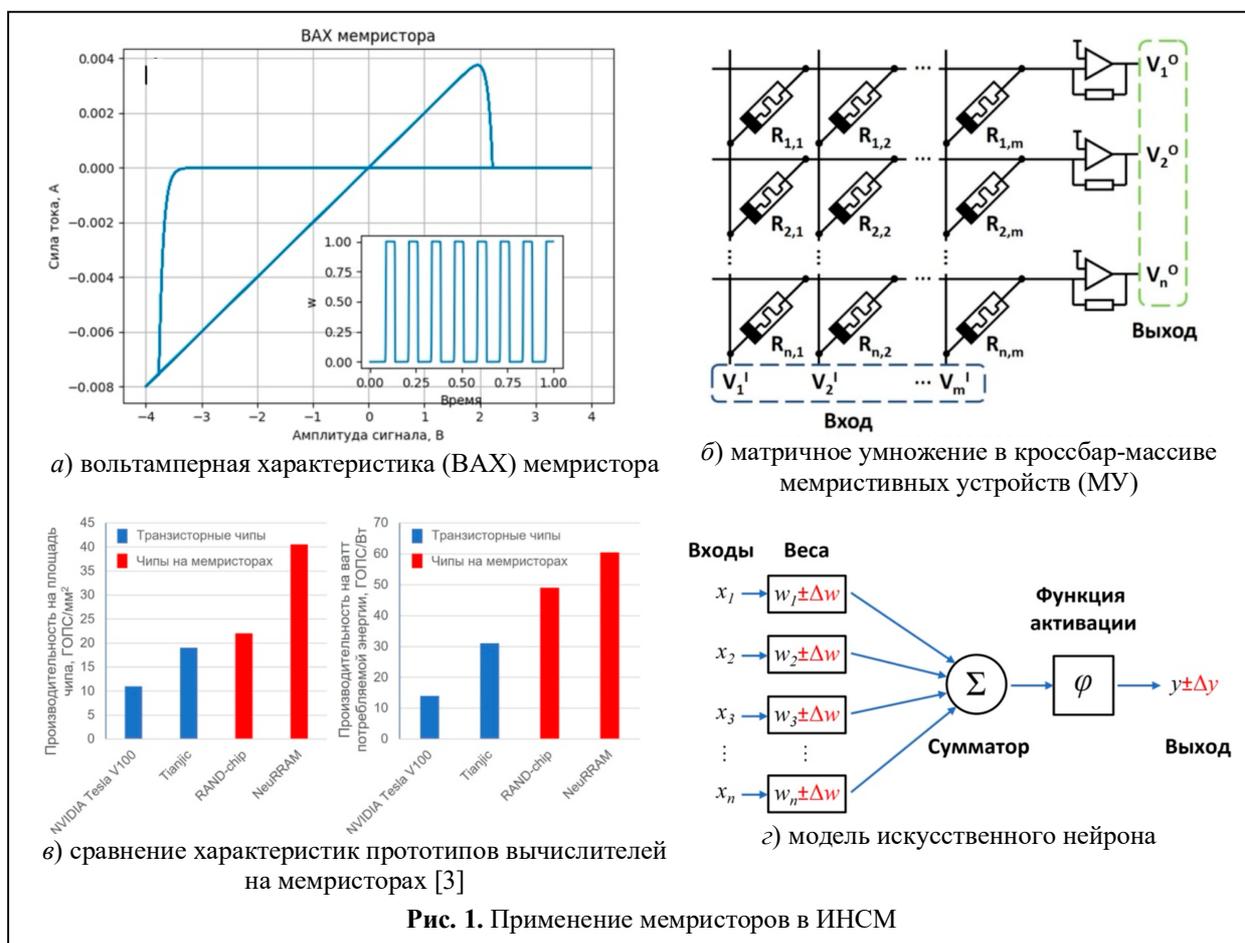
Введение

Мемристор — это пассивный элемент в микроэлектронике, способный изменять своё сопротивление под действием электрического поля и протекавшего через него заряда, а также сохранять его длительное время [1] (рис. 1, а). В искусственных нейронных сетях (ИНС) мемристоры используются для выполнения операции матричного умножения, которое происходит в соответствии с законами Ома и Кирхгофа [2] (рис. 1, б). Это позволяет прототипам ИНС на базе мемристоров (ИНСМ) достигать отличных показателей по производительности на площадь чипа, на ватт потребляемой энергии и другим важным показателям по сравнению с вычислителями на базе классической архитектуры Джона фон Неймана [3] (рис. 1, в). Однако, поскольку вычисления аналоговые, а мемристорам свойственны вариации сопротивлений [4–6], то для того, чтобы создавать нейропроцессоры на базе мемристоров,

нужно уметь оценивать точность работы ИНСМ на этапе проектирования (рис. 1, г), что также регулируется государственными стандартами по надёжности и качеству.

Анализ публикаций показал, что аналитически точность функционирования ИНСМ рассчитать можно лишь для простых демонстрационных случаев. В настоящее время для оценки точности ИНСМ в основном применяется компьютерное моделирование, при котором используются различные модели мемристоров и принципы внесения вариаций в их работу [7–12].

Одними из наиболее распространённых моделей мемристоров являются модели, описывающие связь между приложенным напряжением и током в цепи с мемристором. Эти модели хорошо воспроизводят поведение мемристора, однако при оценке точности ИНСМ имеют ряд недостатков, связанных со сложностью и неоднозначностью в подходах к учёту и



заданию вариаций сопротивлений от цикла к циклу и от устройства к устройству, ресурсоёмкостью процесса моделирования ИНСМ, требующего решения одного или нескольких дифференциальных уравнений для каждого мемристора, что критично для больших ИНСМ состоящих из тысяч или даже миллионов мемристивных устройств (МУ) и др. Кроме того, при таком подходе отсутствует связь с параметрами сигнала задания сопротивления, что добавляет дополнительные трудности при синтезе конкретных значений параметров сигналов после анализа точности. Поэтому оценка точности ИНСМ должна рассматриваться с учётом взаимосвязи между параметрами сигналов задания сопротивления мемристивного устройства — факторами, которые влияют на итоговые сопротивления — и весами, которые влияют на точность ИНСМ.

В статье рассмотрен вопрос оценки точности работы ИНСМ с учётом данной взаимосвязи. В разделе 1 описан подход и предложена методика оценки точности работы. В разделе 2 описана экспериментальная установка, включающая в себя программно-аппаратный комплекс и предназначенная для накопления данных и сравнения модельных оценок с экспериментом. В разделе 3 приведены результаты моделирования и их соответствие экспериментальным данным.

1. Методика

У сигнала, изменяющего сопротивление мемристора, может быть много различных параметров, число комбинаций которых очень большое. Если взять хотя бы три фактора (например, напряжение, длительность, частота) с 10 уровнями, то по формуле комбинаторики размещения с повторениями количество экспериментов будет равно 1000 и это без учёта того, что каждый из этих экспериментов нужно выполнить ещё несколько раз, допустим 1000, что соответствуют 1000000 экспериментов. Провести все эксперименты, покрывающие все комбинации параметров сигналов не-

возможно. Для решения такой задачи подходит теория планирования эксперимента.

В таком случае мемристор рассматривается как черный ящик (рис. 2, а), а модель погрешности мемристивного устройства или веса, будет определяться как некоторая функция, построенная на основе отклика системы на определённые факторы в соответствии планом эксперимента (рис. 2, б). В результате применения теории планирования эксперимента можно получить набор моделей разного уровня структурно-функциональной иерархии для решения различных задач. В рамках данной работы применение теории планирования эксперимента позволит получить набор моделей, основанных на реальных данных о вариациях конкретных мемристивных устройств и позволяющих оценить точность работы ИНСМ с учётом взаимосвязи между параметрами сигналов задания сопротивления и весами ИНСМ.

Использование полученных таким образом моделей погрешности мемристивного устройства и весового коэффициента позволит провести компьютерное моделирование и оценить, какая погрешность будет у ИНСМ после маппирования (переноса весов в кроссбар-массив мемристивных устройств, сопровождающийся сопоставлением шкал весов и сопротивлений с учётом масштабирования) весов и соответственно позволит оценить точность её работы с учётом данных вариаций в реальном устройстве (рис. 2, в). На рис. 2, г схематично показана взаимосвязь между параметрами сигнала задания сопротивления МУ (F_1, F_2, \dots), результатами экспериментальных исследований заданных сопротивлений R и весов W ИНСМ, получаемыми в результате выполнения плана эксперимента и построения модели отклика, и значениями метрики качества инференса (процесса обработки информации) ИНСМ L , получаемыми в результате компьютерного моделирования.

Для оценки точности ИНСМ с помощью описанного подхода можно использовать методику (рис. 2, в), включающую в себя следующие шаги:

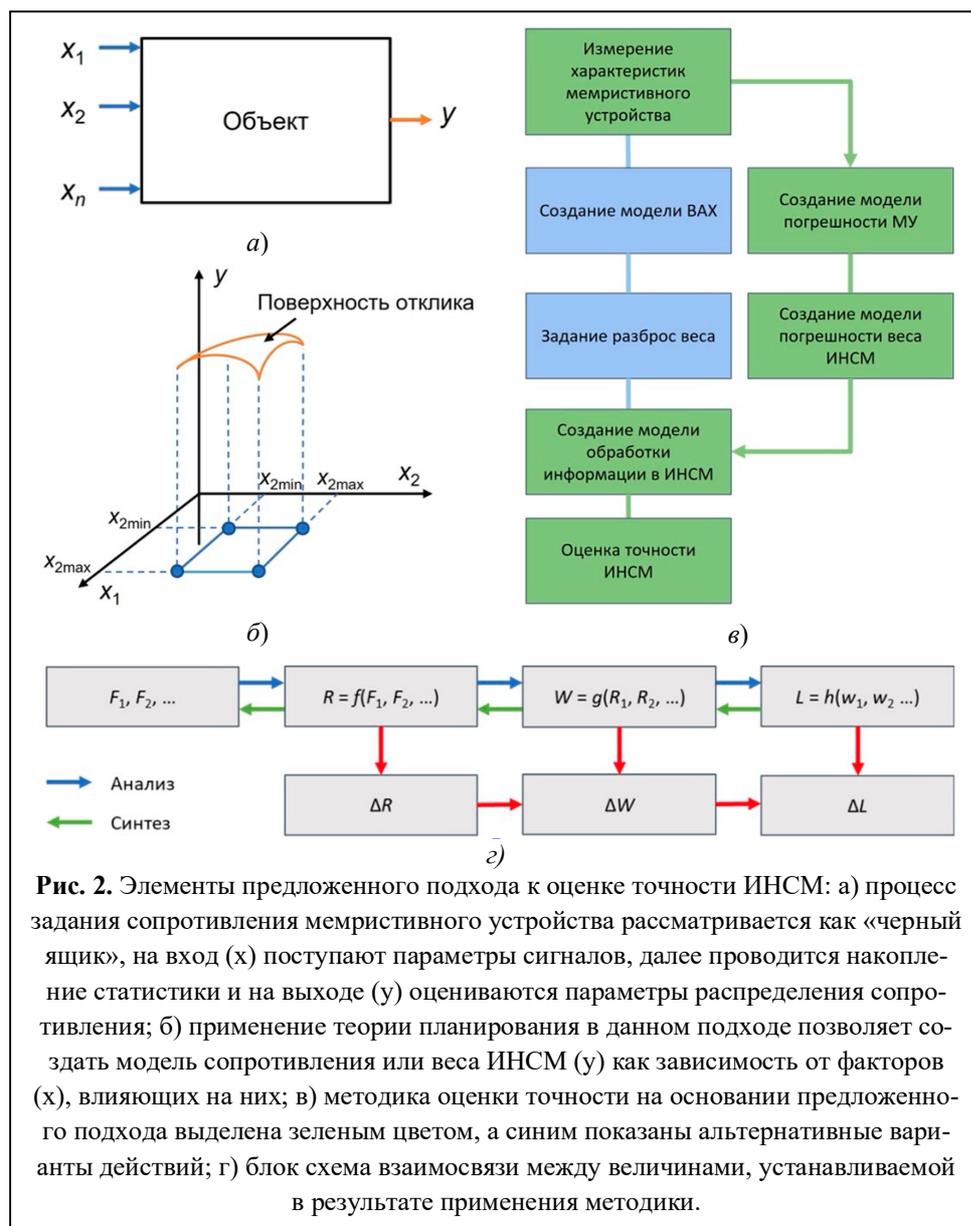


Рис. 2. Элементы предложенного подхода к оценке точности ИНСМ: а) процесс задания сопротивления мемристорного устройства рассматривается как «черный ящик», на вход (x) поступают параметры сигналов, далее проводится накопление статистики и на выходе (y) оцениваются параметры распределения сопротивления; б) применение теории планирования в данном подходе позволяет создать модель сопротивления или веса ИНСМ (y) как зависимость от факторов (x), влияющих на них; в) методика оценки точности на основании предложенного подхода выделена зеленым цветом, а синим показаны альтернативные варианты действий; г) блок схема взаимосвязи между величинами, устанавливаемой в результате применения методики.

1) Создаём модель зависимости сопротивления мемристорного устройства от параметров сигналов его задания.

2) Создаём модель зависимости веса синапса нейрона от сопротивления мемристорного устройства и схемы формирования веса.

Процесс создания моделей 1 и 2 подробно описан в работе [13]. Данные модели создаются на основании результатов измерения характеристик конкретных мемристорных устройств в соответствии с планом эксперимента.

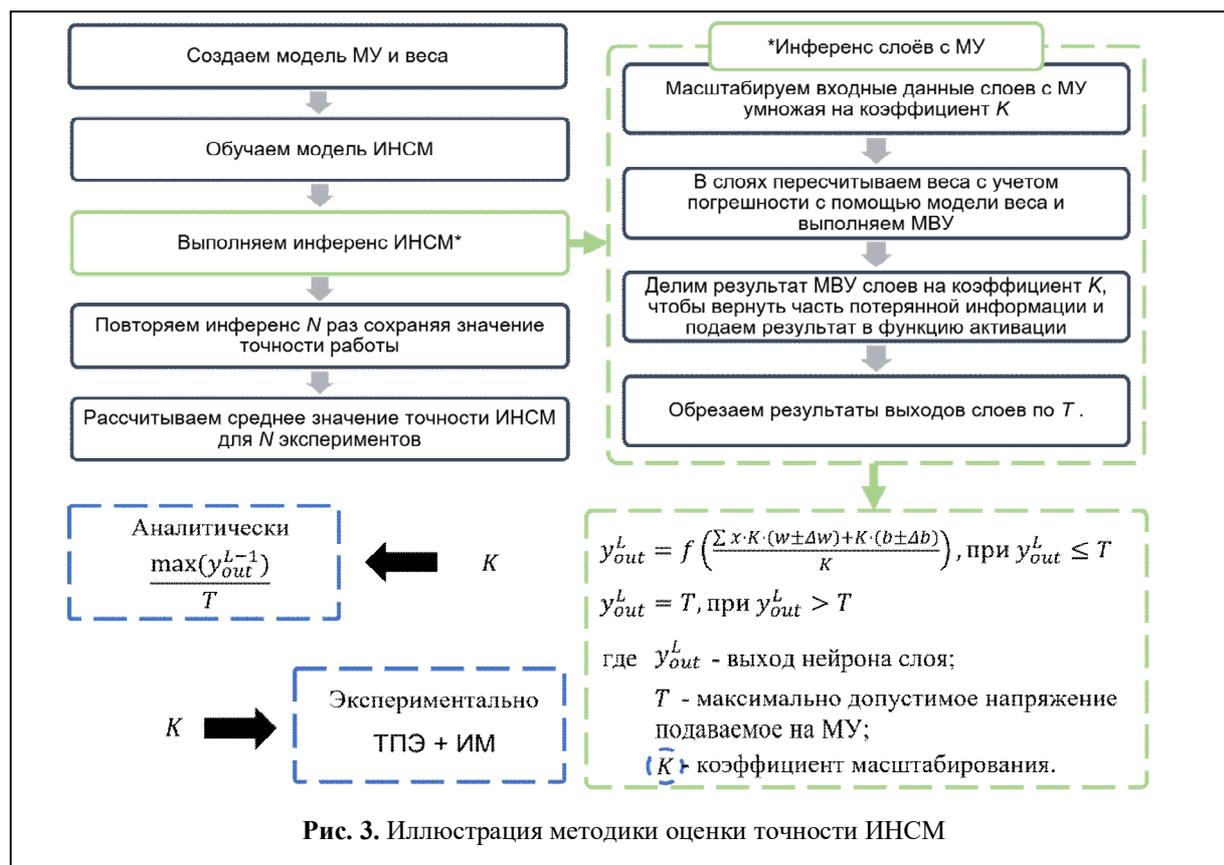
3) Создаём и обучаем решению требуемой задачи модель ИНСМ с учётом конкретной схемы формирования весов синапсов и пара-

метров сигналов кодирования и обработки информации в ней.

4) Выполняем масштабирование весов, в процессе которого для слоёв с МУ генерируем значения весов с вариациями с помощью модели веса. При этом перед выполнением матричного умножения входы слоя необходимо масштабировать, умножив на коэффициент K , а уже после матричного умножения поделить на коэффициент K , чтобы вернуть часть потерянной информации. Затем полученный результат подаётся в функцию активации. После этого у выхода слоя обрезаются значения на уровне T , которые больше допустимых напряжений кроссбар-массива T . Если коэффициент K подобран правильно, то значений превышающих T не будет и потери информации не произойдёт. Если же такие значения будут, то входные данные для следующего слоя уменьшатся до значения T , часть информации будет потеряна и это приведёт к снижению точности ИНСМ.

5) Повторяем пункт 4 N раз, сохраняя значение точности работы.

6) Рассчитываем среднее значение точности ИНСМ для N экспериментов.



Подробная иллюстрация методики представлена на рис. 3.

Данная методика также представляет собой усовершенствованный вариант из работы [13], в который добавлен учёт масштабирования взводных данных слоёв ИНСМ (K) и ограничения максимально допустимых напряжений, подаваемых на вход кроссбара мемристивных устройств (T). Для нахождения масштабирующего коэффициента K может применяться либо аналитический, либо экспериментальный подход.

Аналитически рассчитать масштабирующий коэффициент можно путём выполнения следующих шагов:

1) Берётся обучающая выборка, которая использовалась при обучении ИНСМ, причём входные данные уже должны быть масштабированы для входного слоя в соответствии с особенностями мемристивных устройств;

2) Выполняется имитационное моделирование работы ИНСМ по методу, представленному в предыдущем пункте, используя соответ-

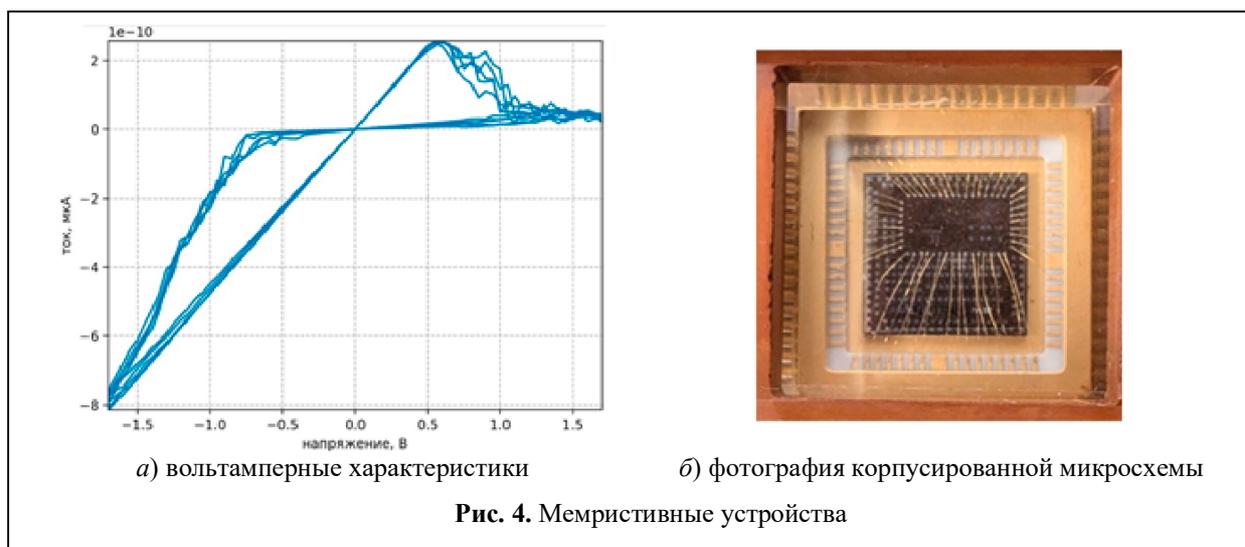
ствующие модели, однако в данном случае нет необходимости оценивать точность. Здесь важным моментом является то, что при моделировании выходы каждого слоя модели ИНСМ необходимо сохранять отдельно, а перед подачей в следующий слой обрезать;

3) В полученных выходных значениях для каждого отдельного слоя находится максимальное число $\max(y_{out}^{L-1})$, которое затем подставляется в формулу (1), для нахождения масштабирующего коэффициента K_L для соответствующего слоя L .

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

где K_L — коэффициент масштабирования выхода L -го слоя; $\max(y_{out}^{L-1})$ — выход L -го слоя ИНСМ; T — максимально допустимое рабочее напряжение, подаваемое на вход кроссбар-массива мемристивных устройств.

Однако данный метод к нахождению масштабирующего коэффициента не всегда работает хорошо. Это связано с тем, что для слиш-



ком низких значений сопротивлений соотношение сигнал/шум уменьшается и шумы, присутствующие в аналоговых вычислениях, начинают оказывать большее влияние на погрешность матричного умножения и соответственно неконтролируемое масштабирование может не только не повысить точность ИНСМ, но и возможно наоборот ещё больше снизить её.

В таком случае масштабирующий коэффициент K может находиться экспериментально с помощью теории планирования эксперимента и включать следующие шаги:

- 1) Выбирается диапазон значений масштабирующего коэффициента K ;
- 2) Для них формируется план эксперимента;
- 3) Выполняется компьютерное моделирование с учётом применения масштабирующего коэффициента, а также ограничений максимально допустимых напряжений T , подаваемых на вход кроссбар-массива для каждого фактора, и вычисляется точность ИНСМ;
- 4) Выбирается тот коэффициент, для которого точность ИНСМ наивысшая.

2. Эксперимент

С целью проверки разработанного общего подхода к оценке точности ИНСМ лабораторией мемристорной нейроэлектроники НОЦ ФТНС ННГУ им. Н.И. Лобачевского представлены мемристивные устройства, реализо-

ванные на основе тонкопленочной структуры «металл-диэлектрик-металл». Исследуемые мемристивные устройства имеют структуру Au/Ta/ZrO₂(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 (рис. 4, а). Кроссбар массив МУ выполнен в форме тестового кристалла (рис. 4, б). 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.

Для работы с мемристивными устройствами в рамках исследования был разработан программно-аппаратный комплекс (ПАК) (рис. 5), 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.

Основные особенности аппаратной части ПАК:

- 1) Модульная система — модуль управления, сигнальный и коммутационный;
- 2) Работа с активными кроссбар-массивами МУ в архитектуре 1Т1М;

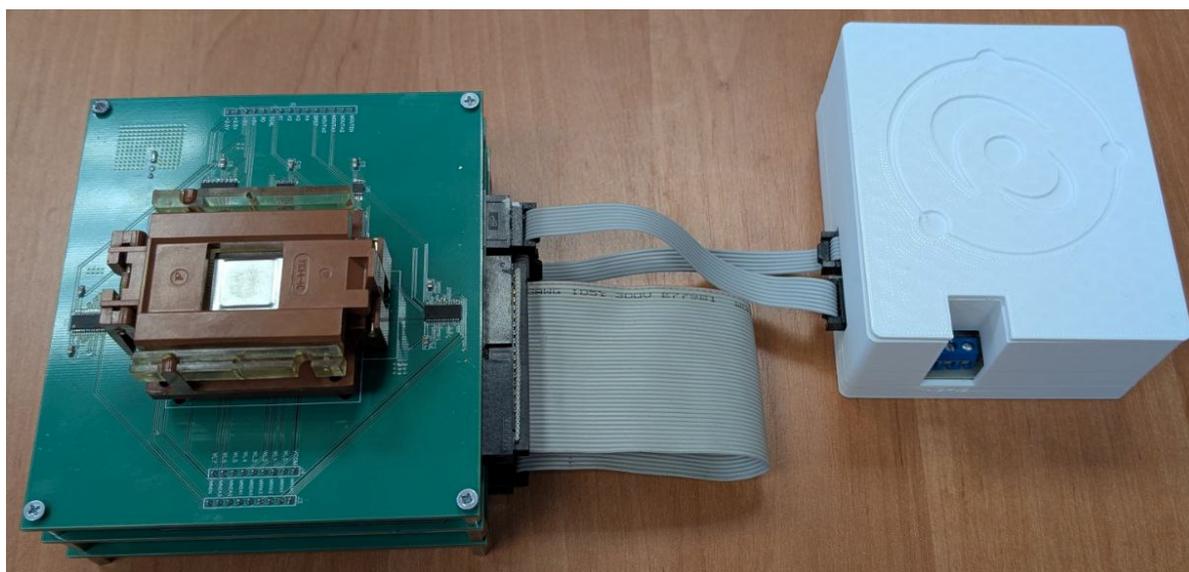


Рис. 5. Фотография аппаратной части разработанного ПАК

3) Питание аппаратной части от лабораторных источников питания (± 12 В);

4) Возможность включения аппаратного ограничителя тока при работе с МУ.

Обмен данными между аппаратной частью и компьютером с установленным пользовательским ПО осуществляется по COM-порту. Разработанное ПО может быть использовано для четыре основных направлений деятельности:

1) Исследование характеристик мемристивных устройств, а именно снятие ВАХ, измерение устойчивости к циклам переключения, снятия синаптической пластичности и других характеристик;

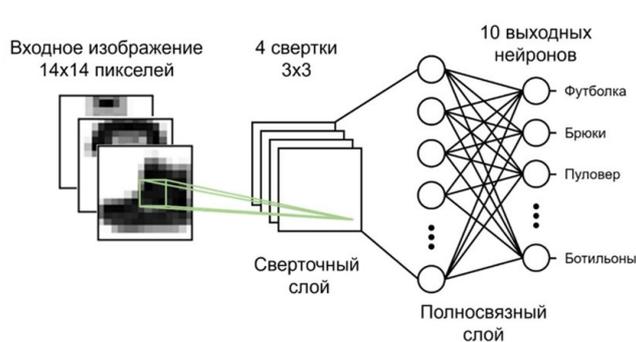
2) Моделирование взаимосвязи между па-

раметрами записи сопротивлений и сопротивлениями МУ и весов синапсов нейронов, как на основе реальных данных, полученных в результате их программирования, так и на основе модельных данных, сгенерированных в LTSpice;

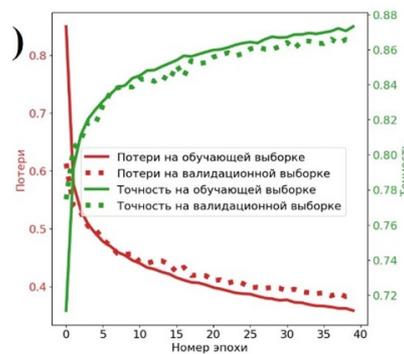
3) Имитационное моделирование ИНСМ с применением моделей или реальных данных о вариациях сопротивлений мемристивных устройств на основе разработанных методов, для оценки точности их работы;

4) Инференса различных архитектур ИНСМ.

Апробация разработанного подхода проиллюстрирована на примере классической задачи машинного обучения по многоклассовой клас-



а) архитектура сети



б) кривые обучения

Рис. 6. ИНС для классификации элементов гардероба на наборе данных Fashion MNIST

сификации на наборе данных Fashion-MNIST (рис. 6). Этот набор данных содержит 70000 изображений элементов гардероба, разделённые 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. Однако для сокращения числа параметров ИНС исходные изображения были с масштабированы до 14 на 14 пикселей.

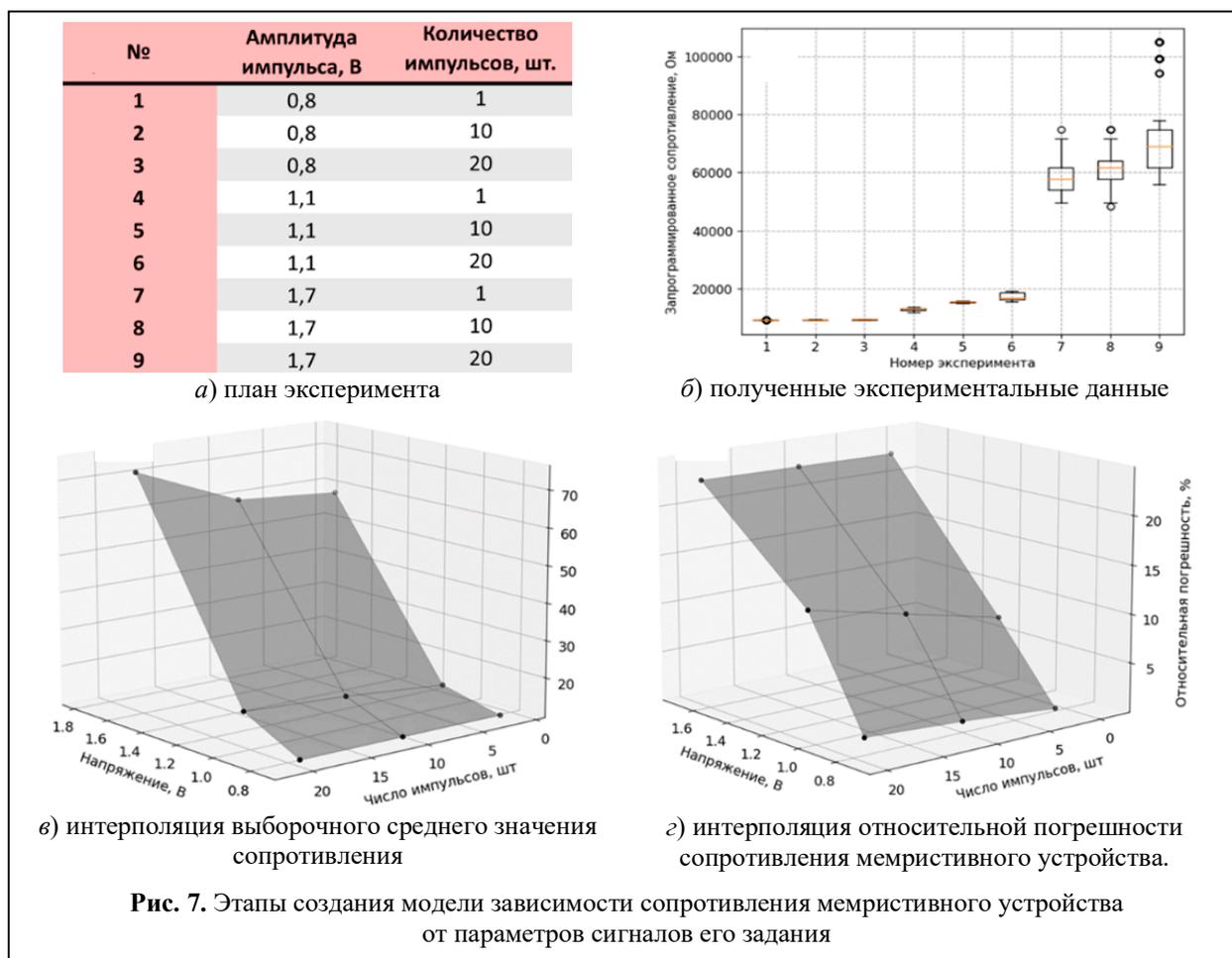
Для решения поставленной задачи применялась свёрточная нейронная сеть. По структуре данная ИНС имеет один свёрточный слой с четырьмя ядрами свёртки размерностью 3 на 3 и функцией активации, линейный выпрямитель (ReLU), а также 10 выходных нейронов в полносвязном слое без функции активации.

Для обучения ИНС использовались 60000

примеров из набора данных, содержащие все виды классов примерно в одинаковых пропорциях, а для тестирования оставшиеся. В итоге точность классификации составила 90% для тестовых данных для программно реализованной модели.

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 точность ИНС немного снизилась до 86,7 %.

Для оценки точности работы данной ИНСМ была разработана модель зависимости сопротивления мемристивного устройства от параметров сигналов его задания с применением



линейной интерполяции на основе экспериментальных данных, полученных в процессе задания сопротивления на кроссбар массиве (рис. 4, б). На рис. 7 представлены основные этапы создания данной модели.

На основе полученной модели мемристивного устройства была создана модель зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса с применением линейной интерполяции, связывающая вес с его погрешностью, основе экспериментальных данных, полученных из модели. На рис. 8 представлены основные этапы создания данной модели.

Для оценки точности работы ИНСМ с учётом погрешности веса выполнялось компьютерное моделирование, которое заключалось в маппировании программной модели ИНСМ 1000 раз и инференсе для всех тестовых данных, при этом весовые коэффициенты на каждой итерации пересчитывались с помощью разработанной модели веса по равномерному закону относительно номинальных квантованных значений весов. Экспериментально измеренные сопротивления для разных комбинаций факторов имеют разные законы распределения, поэтому при моделировании в данном случае использовался равномерный закон распределения, так как он отражает худшие исходы. В таком случае оценка точности будет несколько ниже, чем реальная, однако в данной ситуации ошибиться в худшую сторону лучше, чем получить не корректно высокую оценку точности на модели и низкую точность в устройстве.

3. Результаты и обсуждение

В результате компьютерного моделирования была получена оценка точности ИНСМ в среднем 76,57 %. Для проверки данной оценки была разработана аппаратная реализация ИНСМ на базе созданного ПАК (рис. 5) и проведена экспериментальная оценка точности. Для экспериментальной оценки веса ИНСМ записывались в МУ с параметрами сигнала, используемыми в модели. Экспериментальная точность составила 76,7 %. Если сравнивать точность,

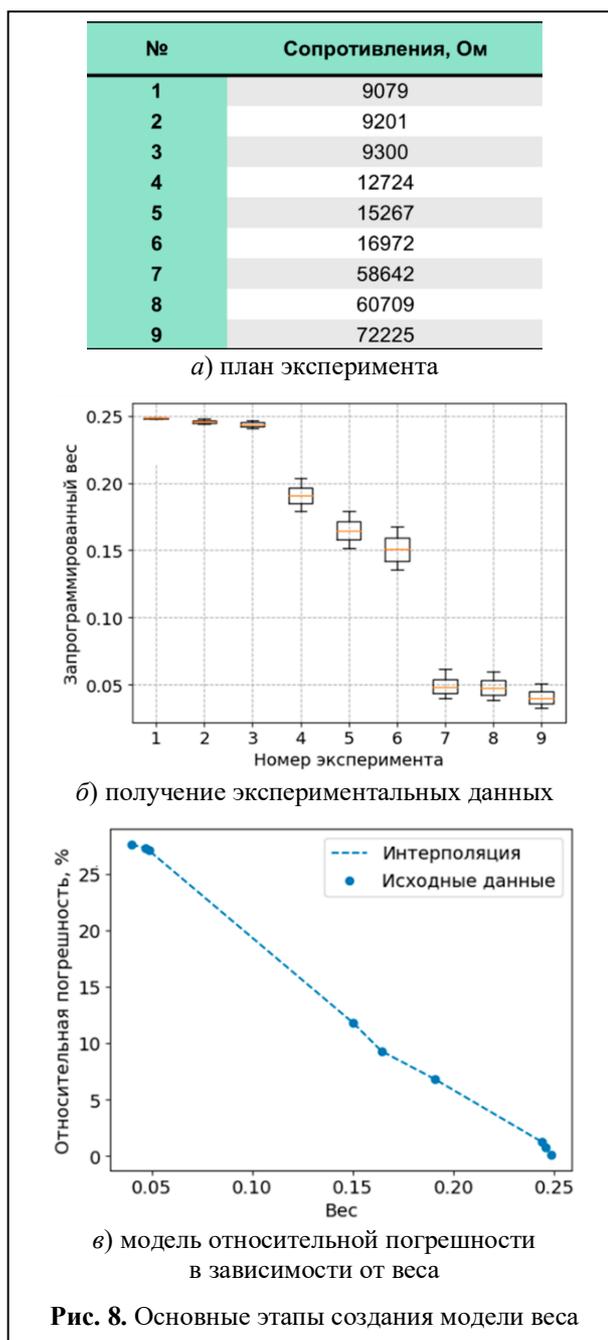


Рис. 8. Основные этапы создания модели веса

оцениваемую с помощью моделирования с реальной точностью, то можно заметить, что разница между ними составляет примерно 0,13 %. Разница в оценке точности между моделью и экспериментом, предположительно, связана с применением равномерного закона распределения, который занижил ожидаемую точность работы ИНСМ, как изначально и предполагалось.

Далее была выполнена оценка вклада погрешностей вычислений вызванных ограниче-

Таблица 1. Результаты имитационного моделирования ИНСМ

Номер эксперимента	Значение масштабирующего коэффициента K	Точность на модели ИНСМ, %
1	0,2	77,36
2	0,3	79,96
3	0,4	79,98
4	0,5	80
5	0,6	76,9

нием максимально допустимых напряжений, подаваемых на вход кроссбара мемристивных устройств. В результате такого исследования было выяснено, что точность ИНСМ при такой реализации составляет в среднем 86,57 %. В связи с этим было решено применить масштабирующий коэффициент K , предложенный в данной работе для уменьшения влияния данного типа погрешностей на итоговую точность.

В данном случае коэффициент K рассчитывался экспериментальным методом так как коэффициент, найденный алгоритмическим методом на модели, привёл к ещё более низкому значению точности 66,7 % при K равном 0,114. Результаты экспериментального расчёта представлены в таблице 1.

На основе приведённых в таблице 1 данных видно, что значение масштабирующего коэффициента 0,5 обеспечивает максимальную точность работы ИНСМ и именно данное зна-

чение нужно использовать в аппаратно реализованной сети. Для проверки данного предположения был выполнен инференс аппаратно реализованной ИНСМ с данным масштабирующим коэффициентом, в результате чего точность классификации составила 80,1 %. Таким образом это наглядно демонстрирует возможности подхода по уменьшению влияния погрешностей, связанных с ограничением максимально допустимых напряжений, подаваемых на вход кроссбар-массива мемристивных устройств. Применение данного подхода позволило увеличить точность с 76,7 % до 80,1 %, то есть на 3,4 %, что составляет более 30% от общего снижения точности.

Подобные эксперименты, результаты которых представлены в сводной таблице 2, были проведены для разных задач и разных архитектур нейронных сетей. В столбце 3 показана точность программной модели ИНСМ без учёта вариаций мемристивных устройств. В столбце 4 показана оценка точности ИНСМ с учётом реальных характеристик мемристивных устройств и сигналов записи сопротивлений. Из этого столбца видно, что вариации мемристоров по-разному влияют на различные архитектуры и структуры ИНСМ: например, для сети №1 точность не ухудшилась, для сетей №3 и 4 ухудшилась не значительно, а для сети №2 снизилась достаточно сильно. С одной стороны, это подтверждает важность проведе-

Таблица 2. Оценка точности для различных моделей ИНСМ

№	Задача	Архитектура ИНСМ	Точность программной модели	Оценка точности ИНСМ с применением разработанной методики	Точность ИНСМ при аппаратной реализации в ПАК
1	Классификация ирисов (Iris Dataset)	Двухслойная сеть прямого распространения	100 %	100 %	100 %
2	Классификация грибов (Secondary Mushroom Dataset)	Трёхслойная сеть прямого распространения	93,9 %	52,7 %	54 %
3	Классификация элементов гардероба (Fashion MNIST Dataset)	Двухслойная свёрточная нейронная сеть	86,75 %	84,9 %	85,6 %
4	Выявление болезни паркинсона (Parkinson Dataset (SST))	Трёхслойная рекуррентная нейронная сеть	72 %	69,7 %	70,5 %

ния предварительной оценки точности работы ИНСМ, поскольку для реального устройства она может сильно отличаться относительно программной модели. С другой стороны, это показывает, что разные наборы весов могут обеспечить разную устойчивость к вариациям сопротивлений в ИНСМ. В столбце 5 показана реальная точность работы ИНСМ при экспериментальной проверке с применением разработанного ПАК. Из столбца 5 действительно видно, что сеть №2 демонстрирует сильное ухудшение точности при её аппаратной реализации. Возможно, её веса не являются оптимальными и в процессе обучения градиентным спуском они были получены для одного из локальных минимумов, такого, что их изменение в пределах погрешности задания веса сильно меняет выходной сигнал сети. В среднем отличие между результатами моделирования и экспериментом составляет не более 3-х процентов. Вероятно, существуют и другие неучтённые факторы, влияющие на точность, которые не рассматривались в рамках данного исследования.

Заключение

Таким образом, в результате выполнения данной работы разработана методика, позволяющая оценить точность работы ИНСМ для конкретных мемристивных устройств и архитектуры сети и заключающаяся в выполнении следующих шагов:

1. Создание модели зависимости сопротивления мемристивного устройства от параметров сигналов его задания;
2. Создание модели зависимости веса синапса нейрона от сопротивления мемристивного устройства и схемы формирования веса;
3. Создание модели ИНСМ и выполнение компьютерного моделирования для оценки точности её работы с учётом выбранных параметров сигналов задания сопротивлений мемристивных устройств и параметров реально заданных сопротивлений, схемы формирования веса и максимально допустимых напряжений на выходе нейронов.

Кроме того, в работе предложена методика расчёта масштабирующего коэффициента K который может находиться как аналитически, так и экспериментально. Данный коэффициент позволяет выполнять масштабирование весов программной модели при их переносе в кроссбар-массивы мемристивных устройств так, чтобы сохранить максимальную точность функционирования, при этом обеспечив безопасные для работы уровни напряжений, кодирующих входные сигналы.

Результаты компьютерного моделирования ИНСМ с использованием предложенной методики были проверены экспериментально. Для этого были созданы и обучены несколько программных моделей ИНС с разной архитектурой, которые затем были аппаратно реализованы в ПАК с применением кроссбар-массивов мемристивных устройств 32×8 1Т1М на основе оксида циркония. Для данных моделей были получены оценки точности с использованием предложенной методики и проведены оценки точности экспериментально. Результаты сравнений показали, что разница в точности между модельной оценкой и реальным устройством для различных архитектур не превышает 3%.

Работа выполнена в рамках научной программы Национального центра физики и математики, направление №9 «Искусственный интеллект и большие данные в технических, промышленных, природных и социальных системах» и в рамках реализации программы стратегического академического лидерства «Приоритет-2030» (Н-432-55_2024-2025, Н-431-55_2024-2025). Мемристивные устройства изготовлены с использованием инфраструктуры Лаборатории мемристивной наноэлектроники (гос. задание № FSWR-2025-0006), Учебного дизайн-центра электроники и Центра коллективного пользования — Научно-образовательного центра «Физика твердотельных наноструктур» Нижегородского государственного университета им. Н.И. Лобачевского.

Литература

1. *Strukov D. B.* et al. The missing memristor found // *Nature*. 2008. Vol. 453. No. 7191. Pp. 80–83.
2. *Amirsoleimani A.* et al. In-memory vector-matrix multiplication in monolithic complementary metal-oxide-semiconductor-memristor integrated circuits: design choices, challenges, and perspectives // *Advanced Intelligent Systems*. 2020. Vol. 2. No. 11. Pp. 2000115.
3. *Мухайлов А.Н.* и др. На пути к реализации высокопроизводительных вычислений в памяти на основе мемристорной электронной компонентной базы // *Физмат*. 2023. Т. 1. No. 1. С. 42–64.
4. *Burr G. W.* et al. Experimental demonstration and tolerancing of a large-scale neural network (165 000 synapses) using phase-change memory as the synaptic weight element // *IEEE Transactions on Electron Devices*. 2015. Vol. 62. No. 11. Pp. 3498–3507.
5. *Chai Z.* et al. The over-reset phenomenon in Ta 2 O 5 RRAM device investigated by the RTN-based defect probing technique // *IEEE Electron Device Letters*. 2018. Vol. 39. No. 7. Pp. 955–958.
6. *Roldán J.B.* et al. Variability in resistive memories // *Advanced Intelligent Systems*. 2023. Vol. 5. No. 6. P. 2200338.
7. *Mehonic A.* et al. Simulation of inference accuracy using realistic RRAM devices // *Frontiers in neuroscience*. 2019. Vol. 13. Pp. 593.
8. *Lammie C.* et al. MemTorch: An open-source simulation framework for memristive deep learning systems // *Neurocomputing*. 2022. Vol. 485. Pp. 124–133.
9. *Giotis C.* et al. Bidirectional volatile signatures of metal-oxide memristors—Part II: Modeling // *IEEE Transactions on Electron Devices*. 2020. Vol. 67. No. 11. Pp. 5166–5173.
10. *Lee Y., Kim K. Lee J.* A compact memristor model based on physics-informed neural networks // *Micro machines*. 2024. Vol. 15. No. 2. Pp. 253.
11. *Alonso F. J.* et al. Memristor variability and stochastic physical properties modeling from a multivariate time series approach // *Chaos, Solitons & Fractals*. 2021. Vol. 143. P. 110461.
12. *Lanza M.* et al. Standards for the characterization of endurance in resistive switching devices // *ACS nano*. 2021. Vol. 15. No. 11. Pp. 17214–17231.
13. *Борданов И. А., Щаников С. А.* Оценка точности работы искусственных нейронных сетей на базе мемристоров с применением моделей на основе данных // *Радиотехнические и телекоммуникационные системы*. 2024. No. 2. С. 61–70.

Поступила 15 января 2025 г.

English

EVALUATION OF THE ACCURACY OF ARTIFICIAL NEURAL NETWORKS BASED ON MEMRISTIVE DEVICES BASED ON THE THEORY OF EXPERIMENTAL PLANNING

Ilya Alekseevich Bordanov — Research engineer, Laboratory for the Development of Artificial Intelligence Systems, Murom Institute¹ (branch) “Vladimir State University named after A.G. and N.G. Stoletovs”.

E-mail: bordanov2011@yandex.ru

Leonid Yaroslavovich Korolev — Research Assistant, Scientific Research Laboratory of Stochastic Multistable Systems, Department of Fundamental and Applied Research, Lobachevsky State University² of Nizhny Novgorod.

E-mail: madimtor@yandex.ru

Sergey Andreevich Shchanikov — PhD, Associate Professor, Scientific Director, Laboratory for the Development of Artificial Intelligence Systems, Murom Institute (branch) “Vladimir State University named after A.G. and N.G. Stoletovs”.

E-mail: seach@inbox.ru

Alexey Nikolaevich Mikhailov — PhD in Physics and Mathematics, Director of the Scientific and Educational Center “Physics of Solid-state Nanostructures” of the Department of Fundamental and Applied Research, Lobachevsky State University² of Nizhny Novgorod.

E-mail: mian@nifti.unn.ru

¹Address: 602264, Russian Federation, Vladimir region, Murom, Orlovskaya St., 23.

²Address: 603022, Russian Federation, Nizhny Novgorod, Gagarin Ave., 23.

Abstract: To date, computer modeling is one of the most common approaches to assessing the accuracy of artificial neural networks based on memristive devices (ANNM), taking into account variations in their resistances. In the process of such modeling, models are usually used that describe the relationship between the

applied voltage and current in a circuit with a memristor. These models reproduce the behavior of the memristor well, however, when evaluating the accuracy of the ANNM, they have a number of disadvantages related to the complexity and ambiguity in approaches to accounting and specifying resistance variations from cycle to cycle and from device to device, the resource intensity of the INSM modeling process, etc. In addition, with this approach, there is no connection with the parameters of the resistance setting signal, which adds additional difficulties in synthesizing specific values of the signal parameters after analyzing the accuracy. In this paper, a new approach to estimating the accuracy of the ANNM at the research design stage is proposed, which consists in the fact that the accuracy of a specific version of the ANNM should be evaluated in relation to the parameters of the resistance setting signal of a memristive device, based on the established relationship between these parameters and the values of resistances and weights actually obtained in the experiment. This approach involves the use of experimental planning theory to create a model of the dependence of the resistance of a memristive device on the parameters of its target signals and a model of the dependence of the weights of the INF synapses on the resistances and the weight formation scheme and their application in computer modeling when evaluating the accuracy of the network.

Keywords: artificial neural networks, memristive devices, computer modeling, computational error, data-based models.

The work was carried out within the framework of the scientific program of the National Center for Physics and Mathematics, direction No. 9 "Artificial intelligence and big data in technical, industrial, natural and social systems" and within the framework of the implementation of the strategic academic leadership program "Priority 2030" (H-432-55_2024-2025, H-431-55_2024-2025). Memristive devices are manufactured using the infrastructure of the Laboratory of Memristor Nanoelectronics (state assignment no. FSWR-2025-0006), the Educational Design Center for Electronics and the Center for Collective Use - the Scientific and Educational Center "Physics solid-state nanostructures" Lobachevsky State University of Nizhny Novgorod.

References

1. *Strukov D. B.* et al. The missing memristor found. *Nature*. 2008. Vol. 453. No. 7191. Pp. 80–83.
2. *Amirsoleimani A.* et al. In-memory vector-matrix multiplication in monolithic complementary metal–oxide–semiconductor-memristor integrated circuits: design choices, challenges, and perspectives. *Advanced Intelligent Systems*. 2020. Vol. 2. No. 11. Pp. 2000115.
3. *Mikhailov A.N.* and others. Towards the implementation of high-performance computing in memory based on a memristor electronic component base // *Fizmat*. 2023. Vol. 1. No. 1. Pp. 42–64.
4. *Burr G. W.* et al. Experimental demonstration and tolerancing of a large-scale neural network (165 000 synapses) using phase-change memory as the synaptic weight element. *IEEE Transactions on Electron Devices*. 2015. Vol. 62. No. 11. Pp. 3498–3507.
5. *Chai Z.* et al. The over-reset phenomenon in Ta₂O₅ RRAM device investigated by the RTN-based defect probing technique. *IEEE Electron Device Letters*. 2018. Vol. 39. No. 7. Pp. 955–958.
6. *Roldán J.B.* et al. Variability in resistive memories. *Advanced Intelligent Systems*. 2023. Vol. 5. No. 6. P. 2200338.
7. *Mehonic A.* et al. Simulation of inference accuracy using realistic RRAM devices. *Frontiers in neuroscience*. 2019. Vol. 13. Pp. 593.
8. *Lammie C.* et al. MemTorch: An open-source simulation framework for memristive deep learning systems. *Neurocomputing*. 2022. Vol. 485. Pp. 124–133.
9. *Giotis C.* et al. Bidirectional volatile signatures of metal-oxide memristors—Part II: Modeling. *IEEE Transactions on Electron Devices*. 2020. Vol. 67. No. 11. Pp. 5166–5173.
10. *Lee Y., Kim K., Lee J.* A compact memristor model based on physics-informed neural networks. *Micromachines*. 2024. Vol. 15. No. 2. Pp. 253.
11. *Alonso F. J.* et al. Memristor variability and stochastic physical properties modeling from a multivariate time series approach. *Chaos, Solitons & Fractals*. 2021. Vol. 143. P. 110461.
12. *Lanza M.* et al. Standards for the characterization of endurance in resistive switching devices. *ACS nano*. 2021. Vol. 15. No. 11. Pp. 17214–17231.
13. *Baranov I.A., Shchanikov S.A.* Evaluation of the accuracy of artificial neural networks based on memristors using data-based models. *Radio engineering and telecommunication systems*. 2024. No. 2. Pp. 61–70.