

МОДЕЛИРОВАНИЕ ОСОБЕННОСТЕЙ РАБОТЫ МЕМРИСТИВНОГО КРОССБАР-МАССИВА В НЕЙРОМОРФНЫХ ЭЛЕКТРОННЫХ МОДУЛЯХ

© 2024 г. А. П. Дудкин, Е. А. Рындин*, Н. В. Андреева

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»
им. В.И.Ульянова (Ленина), Санкт-Петербург, Россия

*E-mail: rynenator@gmail.com

Поступила в редакцию 10.08.2024 г.

После доработки 20.09.2024 г.

Принята к публикации 20.09.2024 г.

Разработаны модель и методика моделирования мемристивных кроссбар-массивов с учетом падений напряжения на межсоединениях, шага перестройки уровней проводимости мемристивных элементов и нелинейности их вольт-амперных характеристик. Получены результаты тестирования импульсной нейронной сети в инференс-режиме в задаче распознавания изображений с применением разработанной методики моделирования с учетом характеристик экспериментально изготовленных мемристивных структур.

Ключевые слова: импульсные нейронные сети, мемристивный кроссбар-массив

DOI: 10.31857/S0544126924060031

1. ВВЕДЕНИЕ

Развитие современной компьютерной техники ориентировано на повышение производительности обработки информации. Одной из основных проблем традиционных вычислительных систем на основе архитектуры фон Неймана является проблема “бутылочного горлышка”, состоящая в пространственном разделении блоков памяти для записи и хранения данных и вычислительных блоков для их обработки [1]. С целью эффективного решения данной проблемы, значительные усилия направлены на разработку альтернативных концепций архитектур, ориентированных на память, в том числе нейроморфных. К архитектурам, ориентированным на память, относятся: *in memory computing* — вычисления в оперативной памяти; *near-memory computing* — вычисления рядом с памятью.

“Вычисления в памяти” позволяют значительно повысить производительность и энергоэффективность аппаратно реализованных вычислительных архитектур, причем максимального снижения энергопотребления и повышения производительности можно достичь посредством аппаратной интеграции кроссбар-массивов мемристоров — новых элементов энергонезависимой памяти, которые, как ожидается, станут ключевыми элементами нейропроцессоров, обеспечивая максимально параллельное выполнение вычислений в памяти для

эффективного решения проблемы “бутылочного горлышка” фон Неймана. Повышение эффективности при использовании матричных вычислений на базе мемристивных кроссбар-массивов достигается за счет уменьшения количества операций, требуемых для аппаратного умножения матриц, по сравнению с тем количеством, которое требуется при использовании как графических ускорителей, так и нейросетевых (тензорных) процессоров. Более того, энергонезависимость мемристоров делает возможным реализацию асинхронного режима работы нейроморфных архитектур. Таким образом, мемристивные кроссбар-массивы обеспечивают простой способ аппаратной реализации матрично-векторных операций, являющихся основными, при аппаратном исполнении импульсных нейронных сетей (ИмНС).

Существуют различные типы кроссбар-массивов, но особое внимание уделяется двум: активным и пассивным. В активных кроссбар-массивах транзистор последовательно включен с мемристором (1T1R ячейка) [2—5], что позволяет существенно снизить токи утечки через соседние ячейки и обеспечивает аккуратное считывание и программирование резистивного состояния мемристоров в массиве. Во время расчетов все транзисторы открыты, во время программирования резистивного состояния (в процессе обучения ИмНС) открыты только транзисторы программируемых мемристоров. В

пассивных кроссбар-массивах (ПКМ) [6–7] вместо транзистора, последовательно с мемристором может использоваться селектор. В режиме чтения, на верхний электрод мемристивной ячейки подается напряжение, соответствующее напряжению чтения структуры, при этом нижний электрод заземлен. По сравнению с активными, пассивные кроссбар-массивы обладают такими преимуществами, как снижение энергопотребления, хорошая (по сравнению с 1T1R структурами) масштабируемость и облегченная 3D интеграция. Однако, при аппаратной интеграции пассивных кроссбар-массивов требуется учитывать влияние на точность вычислений в памяти следующих факторов:

- падения напряжения на интегральных соединительных линиях;
- нелинейности вольт-амперных характеристик реальных мемристорных структур;
- наличие конечного множества резистивных состояний мемристоров, ограничивающих количество возможных синаптических весов ИМНС, которые могут быть записаны в элемент кроссбар-массива.

Программная симуляция особенностей аппаратного исполнения кроссбар-массива в задаче распознавания изображений искусственными нейронными сетями была проведена в работах [8–14]. Данные о полученной точности распознавания с учетом особенностей аппаратной реализации кроссбар-массива представлены в табл. 1. Точность распознавая — это отношение количества корректно определенных изображений в процессе тестирования к общему количеству тестовых изображений, которые были представлены нейронной сети в процессе тестирования.

В табл. 2 приведены результаты по точности распознавания изображений, полученные в работах [15–18] на аппаратно реализованных ПКМ в инференс режиме.

Приведенные в табл. 1, 2 результаты показывают, что эффективность работы нейронных сетей при практической реализации, может снижаться, в худшем случае, более чем на 20% в сравнении с алгоритмическим исполнением. Успешная интеграция OxRAM в нейроморфные системы требует решения задач сопряжения режимов аналоговой

Таблица 1. Данные о точности распознавания изображений без учета и с учетом особенностей аппаратной реализации ПКМ, полученные на программных моделях [8–14]

Работа	База данных	Точность распознавания в инференс режиме программной алгоритмической модели нейросети, %	Точность распознавания в инференс режиме программного макета кроссбар-массива, %	Архитектура	Особенностей кроссбар-массива, которые учтены
2017 г. [8]	MNIST	98	77	Полносвязная сеть 784x300x34	Сопротивление проводов; Разброс проводимости от цикла к циклу;
2021 г. [9]	MNIST	97.5	79.9	Полносвязная сеть, перцептрон 784x100x10	Влияние температурных факторов;
2019 г. [10]	MNIST	93	70	Вероятностная нейронная сеть (PNN) 560 sub-кроссбаров (14x10)	Дискретный диапазон перестройки состояний проводимости; Токи утечки;
2016 г. [11]	MNIST	93	83	Двухслойная сеть 50x50 с применением метода опорных векторов (SVM)	Вариативность устройств; Сопротивление проводов; Электрический шум входного сигнала;
2023 г. [12]	MNIST	93	88	Полносвязная сеть, перцептрон 784x256x10	Дискретный диапазон перестройки состояний проводимости; Вариативность параметров от цикла к циклу;
2023 г. [13]	MNIST (3 класса 14x14 пк.)	93	70	Полносвязная сеть, перцептрон 196x3	Сопротивление проводов; Влияние температурных факторов;
2021 г. [14]	custom	93	80	Кроссбар 128x128	Сопротивление проводов

Таблица 2. Результаты по точности распознавания изображений, полученные в работах [15—18] на аппаратно реализованных ПКМ в инференс режиме

Работа	База данных	Точность распознавания в инференс режиме программной алгоритмической модели нейросети, %	Точность распознавания в инференс режиме аппаратного модуля, %	Разница точности распознавания алгоритмической модели и аппаратного модуля, %	Архитектура
2019 г. [15]	MNIST 8x8	—	—	2	64 × 64 кроссбар-массив
2018 г. [16]	custom	95	79	16	20 × 20 кроссбар массив Мемристивные элементы Pt/Al2O3/TiO2-x/Ti/Pt
2021 г. [17]	custom	100	81	19	20 x 20 кроссбар-массив (2 штуки)
2024 г. [18]	MNIST	96	90.8	5.2	16 x 16 кроссбар-массив Импульсная нейронная сеть

перестройки проводимости мемристивных структур с цифровым управлением, определяя необходимость дискретизации уровней проводимости, а также решения проблем, обусловленных вариативностью параметров и нелинейностью характеристик структур. Поэтому разработка моделей и эффективных средств проектирования мемристивных кроссбар-массивов, учитывающих данные особенности мемристоров, является актуальной и важной задачей.

Пассивный кроссбар-массив — это структура, состоящая из двух систем взаимно ортогональных проводящих линий первого и второго уровней металлизации с мемристивной композицией наноразмерных оксидных пленок между ними. Элементы энергонезависимой резистивной памяти образуются в оксидных пленках на пересечениях проводящих линий первого и второго уровней (рис. 1а).

На строки кроссбара подается вектор входных напряжений V , а со столбцов считывается вектор выходных токов I , представляющий результат матрично-векторного умножения

$$I_j = \sum_{i=1}^m V_i * G_{i,j}, \tag{1}$$

где G — матрица проводимостей (весов) мемристивных элементов (рисунок 1), i и j — номера строки и столбца соответственно.

Такое вычисление при аналоговом подходе выполняется параллельно, с временной сложностью $O(1)$, в отличие от программного исполнения операции векторно-матричного умножения. Обозначение $O(1)$ описывает константную (постоянную) сложность. Временная сложность такого вычисления не зависит от размера кроссбар-массива и является постоянной величиной.

Представленный на рис. 1б принцип функционирования кроссбар-массива использует допущения о нулевом сопротивлении проводников, соединяющих мемристивные элементы, и о линейности вольт-амперных характеристик мемристоров. В этом идеализированном случае приложенные к строкам кроссбар-массива напряжения полностью падают на соответствующих мемристивных элементах, а каждому резистивному состоянию мемристивной структуры соответствует определенное фиксированное значение сопротивления. Выходные токи мемристивного массива, в соответствии с законами Ома и Кирхгоффа, будут определять результат векторно-матричного умножения в соответствии с выражением (1).

В реальных кроссбар-массивах соединительные линии имеют определенное сопротивление, и часть приложенных напряжений будет падать на отрезках соединений. Кроме этого, в отличие от программного исполнения, допускающего условно непрерывный диапазон изменения сопротивлений мемристоров, переключать состояния проводимости реальных структур можно только с определенным шагом, то есть диапазон резистивных состояний будет дискретным. При этом вольт-амперные характеристики (ВАХ) мемристивных элементов нелинейны, в результате чего сопротивление мемристора, соответствующее определенному резистивному состоянию, будет зависеть от приложенного к структуре напряжения. Все перечисленные факторы будут влиять на результат векторно-матричного умножения, особенно при увеличении размерности кроссбар-массива, и должны учитываться в процессе проектирования.

Целью данной работы является разработка методики моделирования мемристивных кроссбар-массивов с учетом сопротивлений соединительных

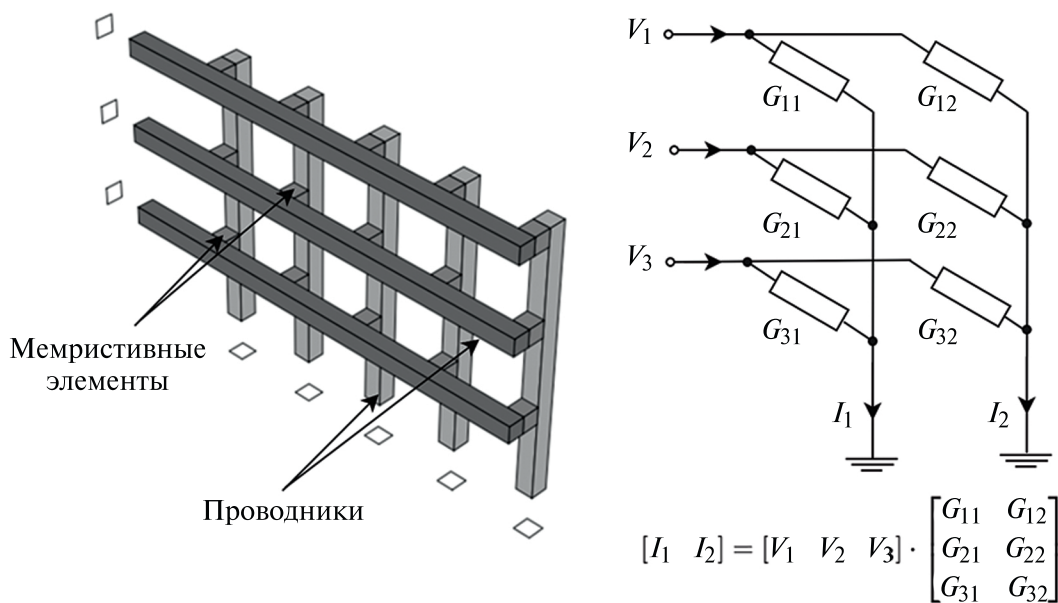


Рис. 1. Мемристивный кроссбар массив: а – структура; б – реализация матрично-векторного умножения.

линий, нелинейности ВАХ и дискретности диапазона резистивных состояний мемристивных элементов.

2. МЕТОДИКА МОДЕЛИРОВАНИЯ МЕМРИСТИВНЫХ КРОССБАР-МАССИВОВ

Для расчета выходных токов кроссбар-массива с учетом сопротивлений соединительных линий, нелинейности ВАХ и дискретности диапазона резистивных состояний мемристивных элементов, при выполнении данного исследования разработана методика моделирования, выраженная в виде итерационного алгоритма, позволяющего находить решение рассматриваемой комплексной задачи методом последовательных приближений. Блок-схема алгоритма представлена на рис. 2.

В качестве входных данных для выполнения алгоритма (блок 1 на рис. 2) используются:

- матрицы весов программно предобученной нейронной сети;
- минимальное R_{MIN} и максимальное R_{MAX} значения сопротивлений мемристивных элементов, измеренные экспериментально;
- сопротивления элементов межсоединений;
- ВАХ мемристоров, соответствующие множеству Ω дискретных резистивных состояний мемристоров;

- целевое значение невязки (параметр σ) – критерий завершения итерационных процедур.

В блоке 2 выполняется перевод матрицы весов программно предобученной нейронной сети с диапазоном значений весов от 0 до 1 (где 0 и 1 соответствуют экспериментально измеренным резистивным состояниям с минимальной и максимальной проводимостью, соответственно) в сопротивления мемристивных элементов кроссбар-массива с учетом заданного диапазона возможных сопротивлений $[R_{MIN}, R_{MAX}]$ в соответствии со следующим алгоритмом:

- 1) перевод весов w_{ij} матрицы, полученной на этапе обучения нейронной сети, в соответствующие сопротивления R_{ij} мемристивных элементов с использованием выражения:

$$R_{ij} = R_{MAX} - (R_{MAX} - R_{MIN})w_{ij}, \quad (2)$$

- 2) корректировка сопротивлений мемристивных элементов посредством приравнивания каждого сопротивления R_{ij} , полученного из выражения (2), к ближайшему из дискретного множества экспериментальных значений Ω .

В блоке 3 на каждой итерации выполняется расчет падений напряжения и токов, протекающих через мемристивные и соединительные элементы кроссбар-массива, в соответствии со схемой, приведенной на рис. 3.

В соответствии с рис. 3, токи, протекающие через строки I_{WLij} , столбцы I_{BLij} и мемристивные

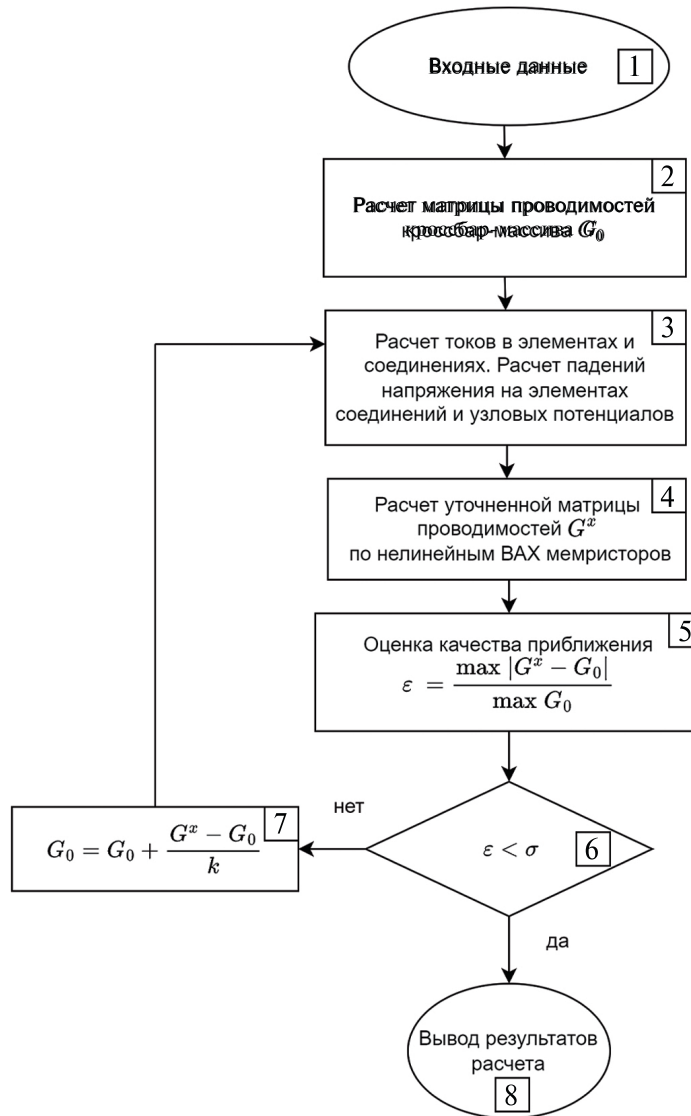


Рис. 2. Блок-схема итерационного алгоритма моделирования мемристивных кроссбар-массивов с учетом сопротивлений соединительных линий, нелинейности ВАХ и дискретности диапазона резистивных состояний мемристивных элементов.

элементы $I_{Di,j}$, можно представить в виде системы уравнений:

$$\begin{aligned}
 I_{WLi,j} - I_{WLi,j+1} - I_{Di,j} &= 0, \quad \text{для } 1 \leq i \leq m; 1 \leq j < n, \\
 I_{WLi,n} - I_{Di,n} &= 0, \quad \text{для } 1 \leq i \leq m, \\
 I_{D1,j} - I_{BL1,j} &= 0, \quad \text{для } 1 \leq j \leq n,
 \end{aligned}
 \quad (3)$$

$$I_{Di,j} + I_{BLi-1,j} - I_{BLi,j} = 0, \quad \text{для } 1 < i \leq m; 1 \leq j \leq n.$$

С учетом закона Ома, токи в уравнениях (3) можно выразить через напряжения следующим образом:

$$\begin{aligned}
 I_{WLi,1} &= G_{WL} (V_i - V_{WLi,1}), \quad \text{для } 1 \leq i \leq m, \\
 I_{WLi,j} &= G_{WL} (V_{WLi,j-1} - V_{WLi,j}), \quad \text{для } 1 \leq i \leq m; 1 < j \leq n, \\
 I_{BLi,j} &= G_{BL} (V_{BLi,j} - V_{BLi+1,j}), \quad \text{для } 1 \leq i < m; 1 \leq j \leq n, \\
 I_{BLm,j} &= G_{BL} (V_{BLm,j}) \quad \text{для } 1 \leq j \leq n, \\
 I_{Di,j} &= G_{ij} (V_{WLi,j} - V_{BLi,j}) \quad \text{для } 1 \leq i \leq m; 1 \leq j \leq n.
 \end{aligned}
 \quad (4)$$

Результирующая система уравнений для расчета токов и напряжений в мемристивном кроссбар-массиве с учетом сопротивлений межсоединений будет иметь вид:

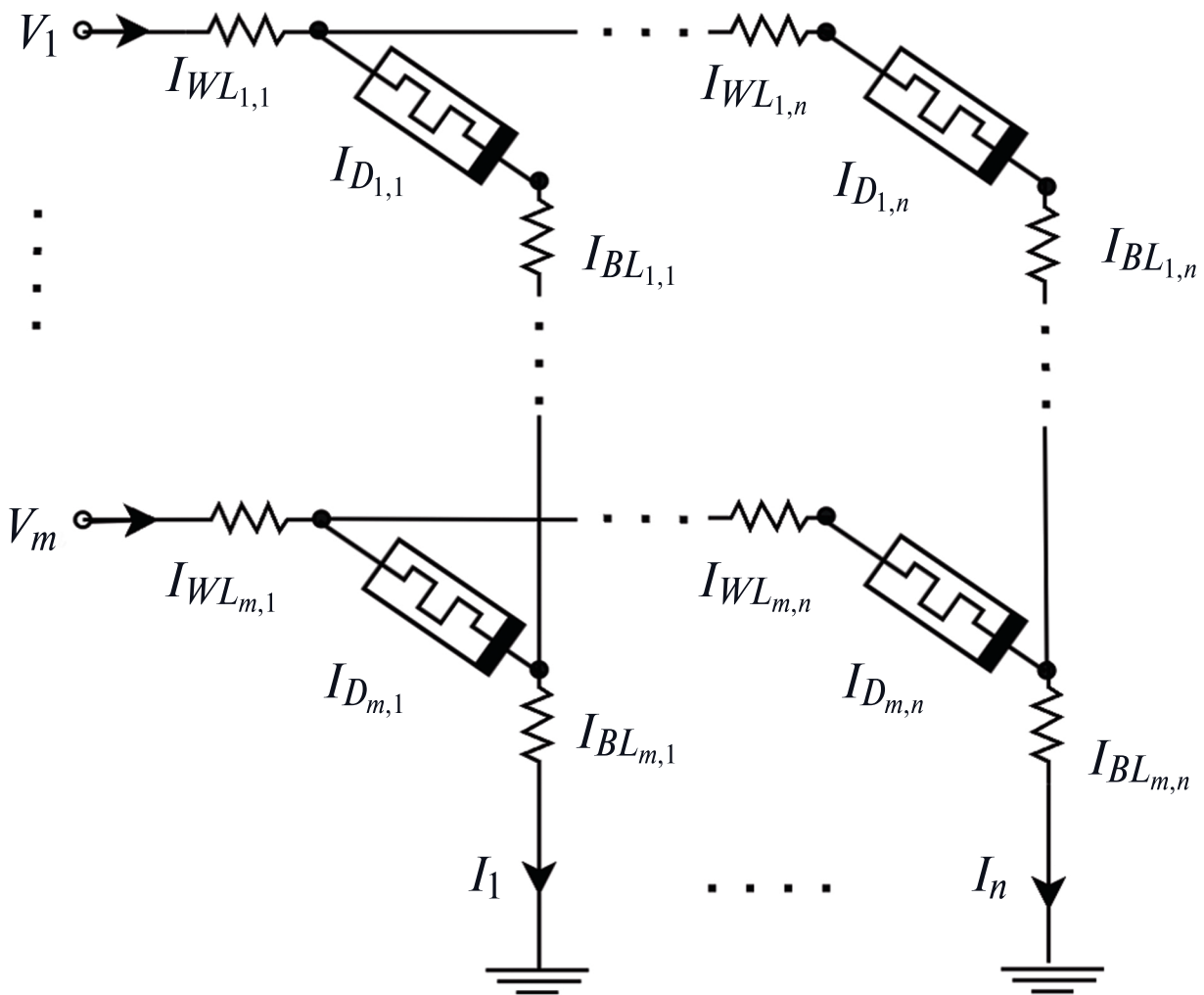


Рис. 3. Схема мемристивного кроссбар-массива.

$$\begin{aligned}
 &(2G_{WL} + G_{i,1}) * V_{WL_{i,1}} - G_{WL} * V_{WL_{i,2}} - G_{i,1} * V_{WL_{i,1}} = G_{WL} * V_i \text{ для } 1 \leq i \leq m, \\
 &(2G_{WL} + G_{i,j}) * V_{WL_{i,j}} - G_{WL} * V_{WL_{i,j-1}} - G_{WL} * V_{WL_{i,j+1}} - G_{ij} * V_{BL_{i,j}} = 0 \\
 &\quad \text{для } 1 \leq i \leq m; 1 < j < n, \\
 &(G_{WL} + G_{i,n}) * V_{WL_{i,n}} - G_{WL} * V_{WL_{i,n-1}} - G_{i,n} * V_{BL_{i,n}} = 0 \text{ для } 1 \leq i \leq m, \\
 &(G_{BL} + G_{1,j}) * V_{BL_{1,j}} - G_{1,j} * V_{WL_{1,j}} - G_{BL} * V_{BL_{2,j}} = 0 \text{ для } 1 \leq j \leq n, \\
 &(2G_{BL} + G_{i,j}) * V_{BL_{i,j}} - G_{i,j} * V_{WL_{i,j}} - G_{BL} * V_{BL_{i-1,j}} - G_{BL} * V_{BL_{i+1,j}} = 0 \\
 &\quad \text{для } 1 \leq i \leq m; 1 < j < n, \\
 &(2G_{BL} + G_{m,j}) * V_{BL_{m,j}} - G_{m,j} * V_{WL_{m,j}} - G_{BL} * V_{BL_{m-1,j}} = 0 \text{ для } 1 \leq j \leq n,
 \end{aligned} \tag{5}$$

где G_{ij} — проводимости мемристивных структур с индексами i, j ; G_{WL} , G_{BL} — проводимости соединительных элементов строк и столбцов соответственно; V_{WL} и V_{BL} — падения напряжения на элементах

соединительных линий; n и m — количество строк и столбцов в кроссбар-массиве.

Система уравнений (5) может быть представлена в матричном виде, как показано на рис. 4.

$$\begin{pmatrix}
 2G_{WL} + G_{1,1} & -G_{WL} & \cdots & -G_{1,1} & 0 & \cdots & 0 \\
 -G_{WL} & 2G_{WL} + G_{1,2} & \cdots & 0 & -G_{1,2} & \cdots & 0 \\
 0 & -G_{WL} & \cdots & 0 & -G_{1,2} & \cdots & 0 \\
 \vdots & & & & & & \\
 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
 \vdots & & & & & & \\
 0 & 0 & \cdots & 0 & 0 & \cdots & -G_{m,n} \\
 -G_{1,1} & 0 & \cdots & G_{BL} + G_{1,1} & 0 & \cdots & 0 \\
 0 & -G_{1,2} & \cdots & 0 & G_{BL} + G_{1,2} & \cdots & 0 \\
 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
 \vdots & & & & & & \\
 0 & 0 & \cdots & 0 & 0 & \cdots & 2G_{BL} + G_{m,n}
 \end{pmatrix}
 \begin{pmatrix}
 [V_{WL}]_{1,1} \\
 [V_{WL}]_{1,2} \\
 \vdots \\
 [V_{BL}]_{1,1} \\
 [V_{BL}]_{1,2} \\
 \vdots \\
 [V_{BL}]_{m,n}
 \end{pmatrix}
 =
 \begin{pmatrix}
 G_{WL} V_1 \\
 \vdots \\
 G_{WL} V_m \\
 \vdots \\
 0 \\
 0 \\
 0 \\
 0 \\
 \vdots \\
 0
 \end{pmatrix}$$

Рис. 4. Система уравнений (5) в матричном виде.

Для решения системы уравнений (5) была разработана программа на языке Python3 с использованием математического пакета SciPy.

В блоке 4 выполняется расчет уточненной матрицы проводимостей мемристивного кроссбар-массива с учетом нелинейности экспериментальных ВАХ мемристивных структур, соответствующих дискретным резистивным состояниям. Экспериментальные ВАХ, необходимые для расчета, хранятся в памяти в виде таблиц значений. Падения напряжения, рассчитанные при выполнении блока 3, используются для извлечения из табличных данных значений токов, соответствующих нелинейным ВАХ, после чего уточненные значения проводимости находятся как отношения токов, протекающих через мемристивные элементы, к падениям напряжения на них.

Для оценки качества текущего приближения, в блоке 5 выполняется расчет невязки ε в соответствии с выражением:

$$\varepsilon = \frac{\max |G^x - G_0|}{\max(G_0)}, \quad (6)$$

где G_0 — матрица проводимостей мемристивных элементов, полученная на предыдущей итерации; G^x — уточненная матрица проводимостей мемристивных элементов, полученная на текущей итерации.

Критерием завершения итерационных процедур является выполнение условия

$$\varepsilon < \sigma, \quad (7)$$

проверка которого выполняется в блоке 6.

Если условие (7) на текущей итерации не выполняется, то в блоке 7 происходит обновление значений матрицы проводимостей G_0 предыдущего приближения в соответствии с выражением:

$$G_0 = G_0 + \frac{G^x - G_0}{k}, \quad (8)$$

где k — параметр, определяющий шаги приращения значений проводимости и используемый для достижения сходимости итерационных процедур.

При выполнении условия (7) осуществляется переход к блоку 8 и вывод значений выходных токов кроссбар-массива.

Предложенные математические модели и методика моделирования мемристивного кроссбар-массива с учетом сопротивлений соединительных линий, нелинейности ВАХ и дискретности диапазона резистивных состояний мемристивных элементов, были реализованы в виде пакета прикладных программ с помощью языка программирования Python3 и фреймворков PyTorch и NumPy.

3. РЕЗУЛЬТАТЫ

Для апробации предложенной методики моделирования мемристивных кроссбар-массивов с учетом сопротивлений соединительных линий, нелинейности ВАХ и дискретности диапазона резистивных состояний мемристивных элементов, были использованы экспериментальные ВАХ мемристивных структур на основе наноразмерной пленки TiOx, приведенные на рис. 5.

В процессе моделирования были рассмотрены высокоомные участки ВАХ экспериментальной структуры, приведенные на рис. 6.

Диапазон напряжений 0–0.7 В соответствует допустимым напряжениям чтения, при которых мемристор не изменяет свое резистивное состояние. ВАХ, показанная на рис. 6 пунктирной линией, получена усреднением экспериментально полученных данных по 20 циклам переключения мемристора.

Как правило, используемое на практике напряжение чтения не превышает 0.2 В. Для такого напряжения чтения среднее значение тока, полученное из экспериментальных данных, составило $1.65 \cdot 10^{-5}$ А при среднеквадратичном отклонении $\sigma = 1.95 \cdot 10^{-6}$ А. Для интеграции в программный симулятор, данные рассматриваемого участка экспериментальных ВАХ аппроксимирова-

лись полиномом третьей степени вида $I(V) = aV^3 + bV + c$. Коэффициенты полинома, рассчитанные с помощью метода наименьших квадратов, имели значения: $a = 2.67 \cdot 10^{-4}$ А/В³; $b = 6.24 \cdot 10^{-5}$ А/В; $c = (1.07 \cdot 10^{-6} + i \cdot 2\sigma)$ А, $i = 0, \dots, N_R$, где N_R – число дискретных резистивных состояний мемристивной структуры. Шаг коэффициента с полинома, определяющий шаг по току между резистивными состояниями, равен 2σ . Такой шаг обусловлен тем, что, исходя из экспериментальных данных, при меньшем шаге однозначное определение текущего резистивного состояния элемента будет не всегда возможно.

Вольт-амперные характеристики, каждая из которых соответствует одному из резистивных состояний мемристивного элемента, полученные с использованием приведенной аппроксимации экспериментальных ВАХ, представлены на рис. 7.

С целью симуляции работы кроссбар-массива была разработана программная модель импульсной нейронной сети (ИмНС) для задачи распознавания изображений рукописных цифр.

Предобученная ИмНС представляла собой полносвязную нейронную сеть из 196 входных и 50 выходных нейронов. Использовалась модель нейронов с адаптивным порогом срабатывания и правилом обучения STDP (правило синаптической пластичности) для синаптических связей (весов) между слоями нейронов. Обучение проводилось с использованием открытой базы данных изображений рукописных цифр MNIST [19]. Полученная матрица весов после программного обучения представлена на рис. 8. Каждый квадрат размером 14×14 элементов на рис. 8 соответствует синаптическим мемристивным элементам одного из пятидесяти выходных нейронов.

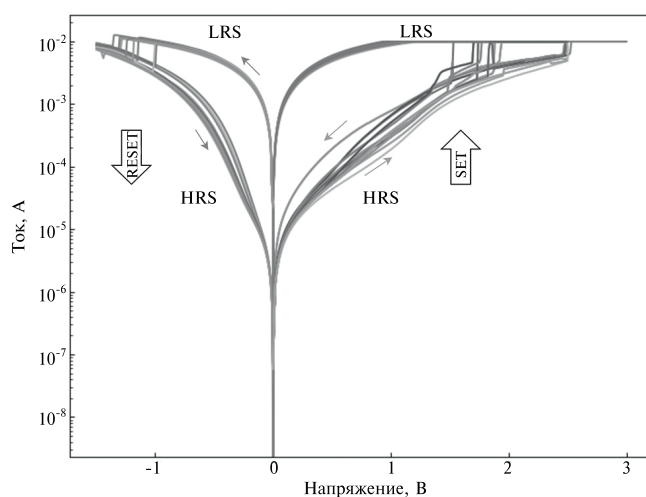


Рис. 5. Экспериментальные ВАХ мемристора на основе наноразмерной пленки TiOx для 20 циклов переключения. HRS и LRS – высокоомное и низкоомное состояния мемристора, соответственно.

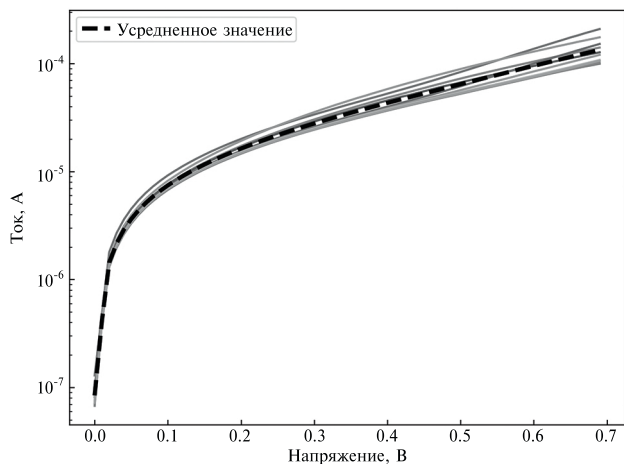


Рис. 6. Высокоомные участки экспериментальных ВАХ мемристора на основе пленки TiOx.

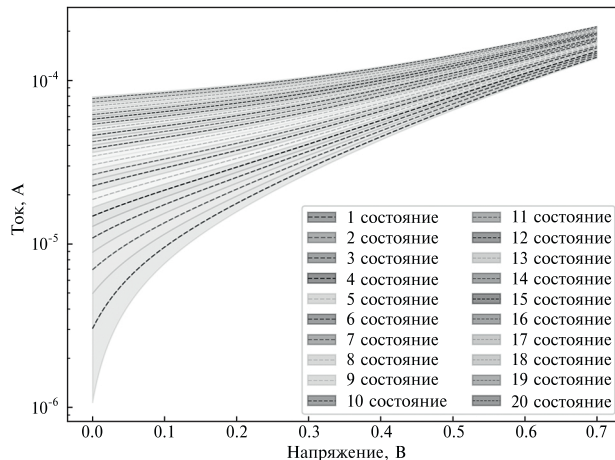


Рис. 7. Семейство ВАХ для различных резистивных состояний, построенное по аппроксимации экспериментальных данных с шагом по току 2σ .

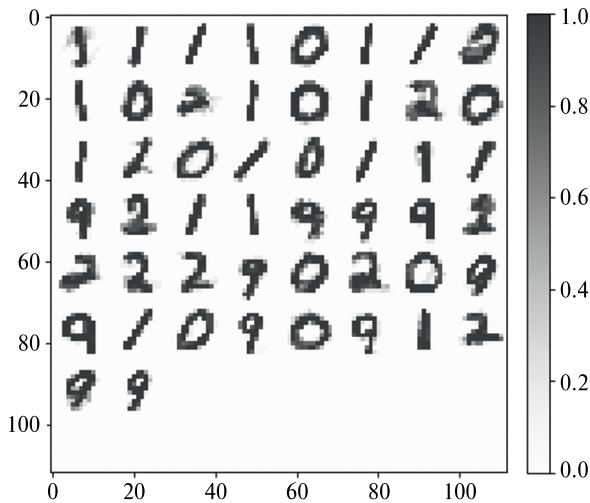


Рис. 8. Матрица весов программно предобученной импульсной нейронной сети.

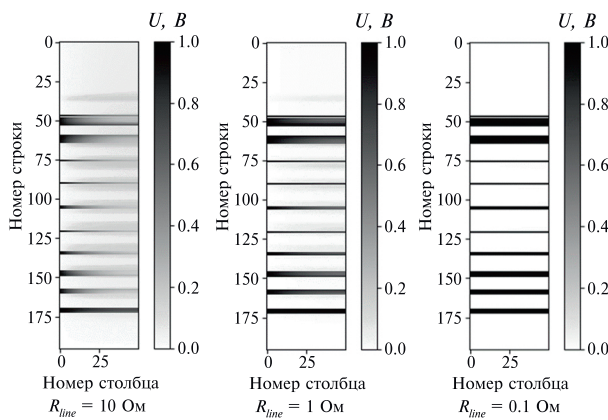


Рис. 9. Падения напряжения на элементах мемристового кроссбар-массива при различных сопротивлениях R_{line} элементов межсоединений.

Полученные весовые коэффициенты элементов кроссбар-массива имели значения в диапазоне от 0 до 1, что в рассматриваемом эксперименте соответствовало диапазону сопротивлений мемристивных элементов 0.1–12 кОм (значению 0 весового коэффициента соответствовало сопротивление 12 кОм, а значению 1 весового коэффициента – сопротивление 0.1 кОм). Данная матрица весов была записана в модель кроссбар-массива в соответствии с описанным выше алгоритмом переноса весов.

Импульсная нейронная сеть тестировалась в инференс-режиме при фиксированной амплитуде импульсов входных напряжений, равной напряжению чтения резистивных состояний мемристоров. Тестирование проводилось на одинаковом количестве тестовых изображений (1000 штук), которые

включали в себя 4 класса (4 цифровых символа). Использовалась модель кроссбар-массива с сопротивлением элементов межсоединений 1 Ом при амплитуде входных сигналов 0.5 В.

Результаты компьютерного моделирования, представленные на рис. 9, показывают, что даже при относительно небольшом размере кроссбар-массива (196×50), сопротивление межсоединений существенно влияет на падение напряжения на мемристивных элементах. На рис. 9 показаны распределения напряжений в кроссбар-массиве при сопротивлениях элементов соединений 10 Ом, 1 Ом, и 0.1 Ом. Максимальные отклонения падений напряжения от идеализированного случая с нулевым сопротивлением элементов межсоединений составило ≈90%, ≈60% и ≈2% соответственно. При увеличении размера кроссбар-массива максимальные отклонения падений напряжения будут пропорционально расти.

Очевидно, что падение напряжения на элементах соединительных линий можно уменьшить увеличением отношения сопротивлений резистивных элементов и элементов межсоединений R_D/R_{line} . Например, как показано на рис. 10, при сопротивлении элементов соединений 10 Ом, увеличение минимального сопротивления мемристивных элементов R_D от 1 кОм до 10 кОм увеличивает соотношение R_D/R_{line} и тем самым уменьшает падение напряжения на межсоединениях.

Визуализация работы итерационного алгоритма (рис. 2) в виде последовательных приближений показана на рис. 11. Высокая скорость сходимости алгоритма (малое число итераций) определяется тем, что, с учетом разброса сопротивлений мемристоров от цикла к циклу, достаточно обеспечить степень приближения результата, соответствующую разбросу ВАХ мемристорных структур в пределах 2σ .

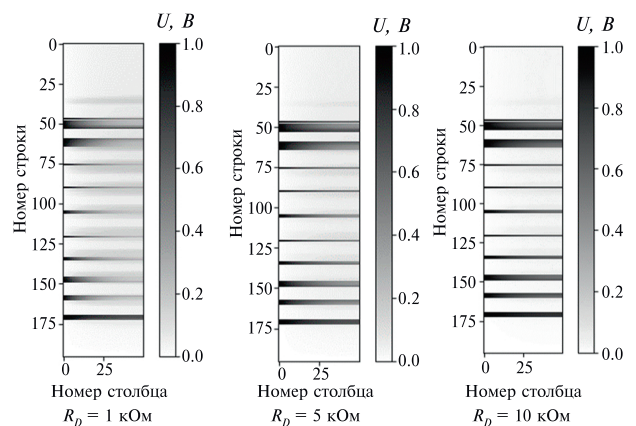


Рис. 10. Падения напряжения в кроссбар-массиве при сопротивлении элементов соединений $R_{line} = 10$ Ом и различных минимальных значениях сопротивлений мемристивных элементов R_D .

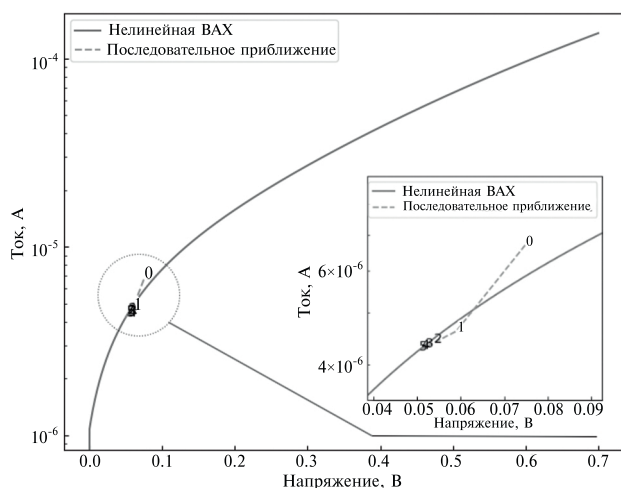


Рис. 11. Пример работы итерационного алгоритма.

Предложенная методика моделирования мемристорных кроссбар-массивов с учетом сопротивлений соединительных линий, нелинейности ВАХ и дискретности диапазона резистивных состояний мемристорных элементов позволяет получать более реалистичную картину распределения токов и напряжений. На рис. 12 представлены токи, протекающие через мемристорные элементы в процессе функционирования мемристорного кроссбар-массива, для идеализированного случая с нулевым сопротивлением соединительных линий и с сопротивлением элементов соединений 1 Ом. Видно, что учет сопротивления соединений приводит к количественным и качественным изменениям результатов, проявляющимся в снижении уровней токов до 50%, а также в появлении отрицательных

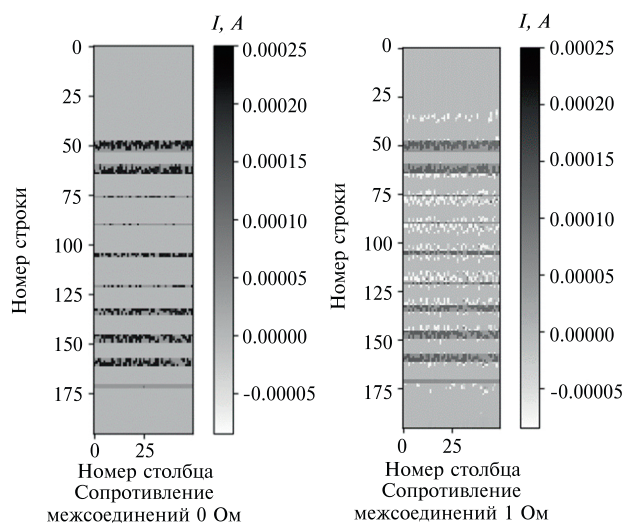


Рис. 12. Сравнение токов, протекающих через элементы мемристорного кроссбар-массива, при нулевом сопротивлении соединительных линий (слева) и сопротивлении элементов соединений 1 Ом (справа).

токов, то есть в смене направлений токов, протекающих через отдельные мемристорные структуры, на противоположные.

Применение предложенного алгоритма для моделирования инференс-режима импульсной нейронной сети с учетом сопротивлений соединительных линий кроссбар-массива, нелинейности ВАХ мемристоров и дискретности резистивных состояний, обеспечило точность распознавания рукописных цифр в среднем 72%. При этом в идеализированном случае, без учета перечисленных факторов, точность распознавания достигла 83%. Таким образом, при учете влияния характеристик кроссбар-массивов и особенностей ВАХ мемристорных элементов эффективность распознавания изображений импульсной нейронной сетью снижается на 9–11% относительно результатов, полученных в идеализированном варианте.

ЗАКЛЮЧЕНИЕ

В работе предложена модель мемристорного кроссбар-массива, позволяющая учитывать влияние особенностей его аппаратной реализации на результаты вычислений в памяти. С использованием экспериментальных ВАХ мемристорных элементов разработана методика учета падения напряжения на элементах соединений, дискретного шага перестройки уровней проводимости мемристорных элементов и нелинейности их вольт-амперных характеристик на работу кроссбар-массива.

Разработанная методика использована для оценки влияния особенностей аппаратной реализации кроссбар-массива на точность распознавания изображений рукописных цифр моделью полностью связанной импульсной нейронной сети.

Полученные результаты тестирования импульсной нейронной сети в инференс-режиме демонстрируют снижение точности распознавания изображений на 9–11% в сравнении с результатами, полученными без учета падений напряжения на элементах соединений, дискретности состояний мемристоров и нелинейности их вольт-амперных характеристик, что свидетельствует о необходимости учета данных факторов и практической важности предложенной методики моделирования мемристорных кроссбар-массивов при проектировании нейроморфных модулей.

ФИНАНСИРОВАНИЕ

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации — государственное задание в области научной деятельности FSEE-2020-0013.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Mutlu O., Ghose S., Gómez-Luna J., and Ausavarungrun R. Processing data where it makes sense: Enabling in-memory computation // Proc. MICPRO, 2019. P. 28–41.
DOI: 10.1016/j.micpro.2019.01.009.
2. Li C., Hu M., Li Y. et al. Analogue signal and image processing with large memristor crossbars // Nature Electronics, 2018. V. 1. P. 52–59.
DOI: 10.1038/s41928-017-0002-z.
3. Li C., Belkin D., Li Y. et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks // Nature Communication, 2018. V. 9. № 1. P. 2385–12385-8.
DOI: 10.1038/s41467-018-04484-2.
4. Li C., Wang Z., Rao M. et al. Long short-term memory networks in memristor crossbar arrays // Nature Machine Intelligence, 2019. V. 1. № 1. P. 49–57.
DOI: 10.1038/s42256-018-0001-4.
5. Hu M., Graves C. E., Li C. et al. Memristor-based analog computation and neural network classification with a dot product engine // Advanced Materials, 2018. V. 30. № 9. P. 1705914-1–1705914-5.
DOI: 10.1002/adma.201705914.
6. Sheridan P., Cai F., Du C. et al. Sparse coding with memristor networks // Nature Nanotechnology, 2017. V. 12. № 8. P. 784–789.
DOI: 10.1038/nnano.2017.83.
7. Du C., Cai F., Zidan M.A. et al. Reservoir computing using dynamic memristors for temporal information processing // Nature Communications, 2017. V. 8. 2204.
<https://doi.org/10.1038/s41467-017-02337-y>.
8. Marinella M.J., Agarwal S., Hsia A., Richter I., Jacobs-Gedrim R., Niroula J., Plimpton S.J., Ipek E., James C.D. Multiscale Co-Design Analysis of Energy, Latency, Area, and Accuracy of a ReRAM Analog Neural Training Accelerator // IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2017. PP. 1–14.
10.1109/JETCAS.2018.2796379.
9. Xu H., Sun Y., Zhu Y., Wang X., Qin G. Statistical Temperature Coefficient Distribution in Analog RRAM Array: Impact on Neuromorphic System and Mitigation Method // <https://arxiv.org/pdf/2105.05534>, 2021.
10. Akhmetov Y., James A.P. Probabilistic Neural Network with Memristive Crossbar Circuits // IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Japan, 2019. P. 1–5.
DOI: 10.1109/ISCAS.2019.8702153.
11. Gu P., Li B., Tang T. et al. Technological exploration of RRAM crossbar array for matrix-vector multiplication // The 20th Asia and South Pacific Design Automation Conference, Chiba, Japan, 2015. P. 106–111.
DOI: 10.1109/ASPDAC.2015.7058989.
12. Xu Q., Wang J., Yuan B. et al. Reliability-Driven Memristive Crossbar Design in Neuromorphic Computing Systems // IEEE Transactions on Automation Science and Engineering, 2023. V. 20. № 1. P. 74–87.
DOI: 10.1109/TASE.2021.3125065.
13. Thomas S. A., Vohra S. K., Kumar R., Sharma R., Das D. M. Analysis of Parasitics on CMOS based Memristor Crossbar Array for Neuromorphic Systems // IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Lansing, MI, USA, 2021. P. 309–312.
DOI: 10.1109/MWSCAS47672.2021.9531867.
14. Zhang L., Borggreve D., Vanselow F., Brederlow R. Impact of Parasitic Wire Resistance on Accuracy and Size of Resistive Crossbars // IEEE International Symposium on Circuits and Systems (ISCAS), Daegu, Korea, 2021. P. 1–5.
DOI: 10.1109/ISCAS51556.2021.9401279.
15. Kim H., Mahmoodi M.R., Nili H. 4K-memristor analog-grade passive crossbar circuit // Nature Communications, 2021. V. 12. 5198.
<https://doi.org/10.1038/s41467-021-25455-0>
16. Bayat F.M., Prezioso M., Chakrabarti B. et al. Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits // Nature Communications, 2018. V. 9. 2331.
<https://doi.org/10.1038/s41467-018-04482-4>
17. Li Y., Ang K.-W. Hardware Implementation of Neuromorphic Computing Using Large-Scale Memristor Crossbar Arrays // Advanced Intelligent Systems, 2020. V. 3. № 1. P. 2000137.
18. Kim S., Hong K., Kim H., Kim M.-H., Choi W. Over-shoot-Suppressed Memristor Array with AlN Oxygen Barrier for Low-Power Operation in the Intelligent Neuromorphic Systems // Advanced Intelligent Systems, 2024. V. 2300797. P. 1–13.
10.1002/aisy.202300797.
19. Deng L. The MNIST database of handwritten digit images for machine learning research // IEEE Signal Processing Magazine, 2012. V. 29. № 6. P. 141–142.

Simulation of Memristive Crossbar Array Electrical Behavior in Neuromorphic Electronic Blocks

A. P. Dudkin, E. A. Ryndin*, N. V. Andreeva

Saint Petersburg Electrotechnical University ETU "LETI", St. Petersburg, Russia

**e-mail: rynenator@gmail.com*

A model and methodology for simulation of memristive crossbar arrays have been developed taking into account voltage drops on interconnections, the step of tuning the conductivity levels of memristive elements and the nonlinearity of their IV characteristics. The results of testing a spiking neural network in the inference mode in the problem of image recognition using the developed methodology of simulation taking into account the characteristics of experimentally manufactured memristor structures have been obtained.

Keywords: spiking neural networks, memristor crossbar array

REFERENCES

1. Mutlu O., Ghose S., Gómez-Luna J., and Ausavarungrun R. Processing data where it makes sense: Enabling in-memory computation // Proc. MICPRO, 2019. P. 28–41, <https://doi.org/10.1016/j.micpro.2019.01.009>.
2. Li C., Hu M., Li Y. et al. Analogue signal and image processing with large memristor crossbars // Nature Electronics. 2018. V. 1. P. 52–59, <https://doi.org/10.1038/s41928-017-0002-z>.
3. Li C., Belkin D., Li Y. et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks // Nature Communication. 2018. V. 9. № 1. P. 2385–12385–8, <https://doi.org/10.1038/s41467-018-04484-2>.
4. Li C., Wang Z., Rao M. et al. Long short-term memory networks in memristor crossbar arrays // Nature Machine Intelligence. 2019. V. 1. № 1. P. 49–57, <https://doi.org/10.1038/s42256-018-0001-4>.
5. Hu M., Graves C.E., Li C. et al. Memristor-based analog computation and neural network classification with a dot product engine // Advanced Materials. 2018. V. 30. № 9. P. 1705914–1–1705914–5, <https://doi.org/10.1002/adma.201705914>.
6. Sheridan P., Cai F., Du C. et al. Sparse coding with memristor networks // Nature Nanotechnology. 2017. V. 12. № 8. P. 784–789, <https://doi.org/10.1038/nnano.2017.83>.
7. Du C., Cai F., Zidan M.A. et al. Reservoir computing using dynamic memristors for temporal information processing // Nature Communications. 2017. V. 8. 2204, <https://doi.org/10.1038/s41467-017-02337-y>.
8. Marinella M.J., Agarwal S., Hsia A., Richter I., Jacobs-Gedrim R., Niroula J., Plimpton S.J., Ipek E., James C.D. Multiscale Co-Design Analysis of Energy, Latency, Area, and Accuracy of a ReRAM Analog Neural Training Accelerator // IEEE Journal on Emerging and Selected Topics in Circuits and Systems. 2017. P. 1–14, <https://doi.org/10.1109/JETCAS.2018.2796379>.
9. Xu H., Sun Y., Zhu Y., Wang X., Qin G. Statistical Temperature Coefficient Distribution in Analog RRAM Array: Impact on Neuromorphic System and Mitigation Method. 2021, <https://arxiv.org/pdf/2105.05534>.
10. Akhmetov Y., James A.P. Probabilistic Neural Network with Memristive Crossbar Circuits // IEEE International Symposium on Circuits and Systems (ISCAS). Sapporo, Japan. 2019. P. 1–5, <https://doi.org/10.1109/ISCAS.2019.8702153>.
11. Gu P., Li B., Tang T. et al. Technological exploration of RRAM crossbar array for matrix-vector multiplication // The 20th Asia and South Pacific Design Automation Conference. Chiba, Japan. 2015. P. 106–111, <https://doi.org/10.1109/ASPDAC.2015.7058989>.
12. Xu Q., Wang J., Yuan B. et al. Reliability-Driven Memristive Crossbar Design in Neuromorphic Computing Systems // IEEE Transactions on Automation Science and Engineering. 2023. V. 20. № 1. P. 74–87, <https://doi.org/10.1109/TASE.2021.3125065>.
13. Thomas S.A., Vohra S.K., Kumar R., Sharma R., Das D.M. Analysis of Parasitics on CMOS based Memristor Crossbar Array for Neuromorphic Systems // IEEE International Midwest Symposium on Circuits and Systems (MWSCAS). Lansing, MI. USA. 2021. P. 309–312, <https://doi.org/10.1109/MWSCAS47672.2021.9531867>.
14. Zhang L., Borggreve D., Vanselow F., Brederlow R. Impact of Parasitic Wire Resistance on Accuracy and Size of Resistive Crossbars // IEEE International Symposium on Circuits and Systems (ISCAS). Daegu, Korea. 2021. P. 1–5, <https://doi.org/10.1109/ISCAS51556.2021.9401279>.

15. Kim H., Mahmoodi M.R., Nili H. 4K-memristor analog-grade passive crossbar circuit // Nature Communications. 2021. V. 12. 5198, <https://doi.org/10.1038/s41467-021-25455-0>
16. Bayat F.M., Prezioso M., Chakrabarti B. et al. Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits // Nature Communications. 2018. V. 9. 2331, <https://doi.org/10.1038/s41467-018-04482-4>
17. Li Y., Ang K.-W. Hardware Implementation of Neuromorphic Computing Using Large-Scale Memristor Crossbar Arrays // Advanced Intelligent Systems. 2020. V. 3. № 1. P. 2000137.
18. Kim S., Hong K., Kim H., Kim M.-H., Choi W. Overshoot-Suppressed Memristor Array with AlN Oxygen Barrier for Low-Power Operation in the Intelligent Neuromorphic Systems // Advanced Intelligent Systems. 2024. V. 2300797. P. 1—13, <https://doi.org/10.1002/aisy.202300797>.
19. Deng L. The MNIST database of handwritten digit images for machine learning research // IEEE Signal Processing Magazine. 2012. V. 29. № 6. P. 141—142.