

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

На правах рукописи

БУСЫГИН АЛЕКСАНДР НИКОЛАЕВИЧ

**ЗАПОМИНАЮЩАЯ МАТРИЦА НА ОСНОВЕ НАНОРАЗМЕРНОГО
КОМБИНИРОВАННОГО МЕМРИСТОРНО-ДИОДНОГО КРОССБАРА
ДЛЯ БИОМОРФНОГО НЕЙРОПРОЦЕССОРА**

Специальность 2.2.2 – электронная компонентная база
микро- и наноэлектроники, квантовых устройств

Диссертации на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
доктор физико-математических наук
Удовиченко Сергей Юрьевич

Тюмень 2021

Оглавление

| | |
|---|-----------|
| ВВЕДЕНИЕ | 4 |
| ГЛАВА 1. Запоминающие матрицы для хранения информации и обработки входных данных путем взвешивания и суммирования сигналов..... | 12 |
| 1.1. Выводы к главе 1. | 31 |
| ГЛАВА 2. Запоминающая матрица для аппаратной нейросети биоморфного нейропроцессора. | 33 |
| 2.1. Физические основы создания ячейки сверхбольшой запоминающей матрицы | 33 |
| 2.2. Электрическая схема ячейки на основе комплементарных мемристоров. | 34 |
| 2.3. Электрические схемы планарной двухслойной и 3D матрицы на основе интеграции мемристорных ячеек. | 35 |
| 2.4. Топология комбинированного мемристорно-диодного кроссбара... .. | 39 |
| 2.5. Топология 3D матрицы на основе интеграции двухслойного мемристорно-диодного кроссбара..... | 41 |
| 2.6. Выводы к главе 2. | 43 |
| ГЛАВА 3. SPICE - моделирование работы сверхбольшой запоминающей матрицы..... | 44 |
| 3.1. Физико-математическая модель работы ячейки матрицы. | 47 |
| 3.2. Физико-математическая модель работы запоминающей матрицы. . | 54 |
| 3.3. Численное моделирование процессов обработки сигналов при взвешивании импульсов напряжения и сложении токов ячеек в запоминающей матрице и сравнение с экспериментальными данными. .. | 55 |
| 3.4. Сравнение результатов численного моделирования работы запоминающей матрицы с известными аналогами | 59 |

| | |
|--|-----------|
| 3.5. Выводы к главе 3 | 61 |
| ГЛАВА 4. Исследование работоспособности запоминающей матрицы в составе аппаратной импульсной нейросети..... | 63 |
| 4.1. Использование запоминающей матрицы в качестве матрицы синапсов аппаратной импульсной нейросети..... | 63 |
| 4.2. Физико-математическая модель ассоциативного самообучения синапсов в запоминающей матрице на основе локального правила обучения LTP..... | 67 |
| 4.3. Физико-математическая модель ассоциативного самообучения синапсов в запоминающей матрице на основе локального правила обучения STDP..... | 70 |
| 4.4. Численное моделирование процесса ассоциативного самообучения запоминающей матрицы в составе аппаратной импульсной нейросети с использованием экспериментальных ВАХ ячеек..... | 71 |
| 4.5. Выводы к главе 4 | 77 |
| ЗАКЛЮЧЕНИЕ..... | 79 |
| СПИСОК ЛИТЕРАТУРЫ..... | 81 |
| ПРИЛОЖЕНИЯ | 92 |

ВВЕДЕНИЕ

Актуальность темы.

Нейронные сети, построенные на простых нейронах и используемые в информационных технологиях, предназначены для аппаратного ускорения расчетов и обеспечивают работу компьютерного зрения, машинного обучения и других систем со слабым искусственным интеллектом. Принятие решения в таких нейросетях происходит в результате выбора наиболее правдоподобного решения на основе ранее заложенных ассоциаций.

В [1] представлена разработка биоморфного нейропроцессора на основе мемристорно-диодного кроссбара, реализующего аппаратную биоморфную импульсную нейросеть с большим числом нейронов для традиционных задач обработки информации, в том числе распознавания паттернов в видео и аудио информации, а также для воспроизведения работы кортикальной колонки мозга или её фрагмента. В качестве ключевых узлов аппаратной части нейропроцессора используются сверхбольшие запоминающая и логическая матрицы, представляющие собой массив синапсов и задающие вес и маршрут связи между нейронами соответственно. Указанные матрицы должны быть сверхбольшими потому, что каждый нейрон в сети может обладать большим количеством синаптических связей. Аналогов разработанного биоморфного нейропроцессора нет. Уникальность нейропроцессора состоит в том, что он построен на основе оригинальной электрической биоморфной модели нейрона и является биоморфным еще и с точки зрения выполнения функций биоморфной нейросети, созданной на основе оригинальной программной биоморфной модели нейрона.

В отличие от нейропроцессоров на простых нейронах биоморфный нейропроцессор дает возможность принимать решения не только на основе заранее заложенных ассоциаций, но и на основе новых ассоциаций (нового знания), формируемых в процессе обработки сигналов в динамично меняющихся условиях. Осмысливание полученных новых ассоциаций при

совершенствовании оригинальной биоморфной нейросети [2] процессора обеспечит переход от слабого к сильному искусственному интеллекту.

В настоящее время существующие аппаратные средства в виде запоминающих матриц с мемристорным кроссбаром [3-7] выполняют узкоспециальные функции, в основном матричное умножение, которое представляет собой взвешенное суммирование входных импульсов напряжения. Аппаратные реализации ассоциативного самообучения и формирование новой ассоциации в импульсной нейросети с дискретными мемристорными синапсами представлены в [8-12]. Однако, предложенные электрические цепи аппаратной реализации ассоциативной памяти не могут быть использованы для построения сверхбольшой аппаратной нейросети. Причиной является отсутствие интеграции мемристоров в кроссбары и наличие в схемах нейронов и синапсов большого числа активных электронных элементов с высоким энергопотреблением. Хотя аппаратные средства [6, 7] содержат в себе интегрированные мемристорные кроссбары, ассоциативное обучение в них является отложенным из-за программного расчета подстройки весов в периферийной системе после прохождения информации. Процедура обновления весов приводит к потере информации, так как во время этого входная информация игнорируется.

Цель работы: создание сверхбольшой запоминающей матрицы биоморфного нейропроцессора, являющейся массивом синапсов искусственных нейронов на основе комбинированного мемристорно-диодного кроссбара –нового компонента наноэлектроники.

Задачи исследования:

- Разработать физические основы создания ячейки сверхбольшой запоминающей матрицы.
- разработать электрические схемы, топологии ячейки и сверхбольшой запоминающей матрицы (1000x1000 ячеек) в планарной и 3D геометрии с высокими интеграцией элементов и энергоэффективностью;

- построить физико-математические модели ячейки и работы запоминающей матрицы в процессах обработки сигналов и ассоциативного самообучения;
- провести численное моделирование работы отдельной сверхбольшой запоминающей матрицы и сравнить с экспериментальными данными;
- провести численное моделирование матрицы малого размера в составе аппаратной импульсной нейросети и сравнить с экспериментальными данными.

Научная новизна.

1. Разработаны физические основы создания ячейки сверхбольшой запоминающей матрицы: необходимость высокого входного сопротивления, которое достигается при комплементарном включении двух мемристоров и обеспечивает малое энергопотребление; выбор в качестве селективного элемента диода Зенера, обеспечивающего существенно большую интеграцию элементов и меньшее энергопотребление по сравнению с активным селектором – транзистором.
2. Созданы электрическая схема и физико-математическая модель работы ячейки сверхбольшой запоминающей матрицы на основе комбинированного мемристорно-диодного кроссбара, ячейка которого содержит комплементарно включенных мемристоров и диода Зенера.
3. Показано, что применение диода Зенера в качестве селективного элемента, подключенного электрически последовательно к мемристору, позволяет увеличить размер кроссбара за счет уменьшения паразитных токов между ячейками, а также уменьшить деградацию выходного сигнала при суммировании входных импульсов напряжения в сверхбольшой матрице.
4. Предложена топология на кристалле как в планарном исполнении, так и с применением 3D интеграции на основе нового компонента нанoeлектроники – комбинированного мемристорно-диодного кроссбара.

5. Разработаны электрическая схема и физико-математические модели процессов обработки сигналов в запоминающей матрице и её ассоциативного самообучения по правилам долговременной пластичности и пластичности, зависимой от времени импульса. Модели самообучения обеспечивают специфичную реализацию этих правил, связанную с оригинальностью мемристорно-диодной ячейки.
6. С помощью численного моделирования при использовании измеренных характеристик изготовленных мемристоров продемонстрированы способности отдельной запоминающей матрицы к обработке сигналов по взвешиванию импульсов напряжения и сложению токов ячеек, а также к ассоциативному самообучению запоминающей матрицы в составе аппаратной импульсной нейросети. Результаты численного моделирования хорошо согласуются с экспериментальными данными, что подтверждает корректность разработанных физико-математических моделей процессов, протекающих в запоминающей матрице, и дает основание для использования этих моделей при проектировании матриц сверхбольшого размера.

Научная и практическая значимость.

Разработанная запоминающая матрица на основе мемристорно-диодного кроссбара является одной из важнейших частей нейропроцессора, который аппаратно воспроизводит работу импульсной (биоморфной) нейросети, способной к ассоциативному самообучению. Запоминающая матрица в составе аппаратной биоморфной нейросети процессора генерирует новые ассоциации по биологически подобному механизму. Это позволяет говорить о формировании процессора нового поколения, который качественно отличается от существующих нейропроцессоров для компьютерного зрения, машинного обучения и других систем со слабым искусственным интеллектом.

Полученные результаты работы важны при переходе от существующих самообучающихся аппаратных импульсных нейросетей на дискретных

мемристорах к сверхбольшим нейросетям с синапсами, интегрированными в мемристорно-диодный кроссбар.

С помощью запоминающей матрицы, встроенной в аппаратную импульсную нейросеть, поступление новой неизвестной информации отождествляется с наблюдаемой генерацией новых ассоциаций в биоморфном нейропроцессоре. Переход от слабого к сильному искусственному интеллекту возможен при совершенствовании биоморфной нейросети [13], которая позволит осмысливать эти ассоциации.

Предложенные электрические схемы и топологии сверхбольшой запоминающей матрицы совместимы с имеющимся оборудованием, применяемым для создания КМОП микросхем и защищены патентами РФ на изобретение №2649657 и №2697623.

Разработанная запоминающая матрица войдет в состав нейропроцессора при проектировании его экспериментального образца совместно с АО НИИ Молекулярной Электроники (НИИМЭ) – головного предприятия приоритетного технологического направления «Электронные технологии» РФ и его изготовлению на ПАО «Микрон», г. Москва, Зеленоград. Ранее было достигнуто соглашение о сотрудничестве между ТюмГУ и компанией АО «Ангстрем-Т» по созданию экспериментального образца аппаратного устройства нейропроцессора.

На защиту выносятся:

1. Разработанные физические основы, физико-математическая модель работы и электрическая схема ячейки сверхбольшой матрицы на основе комплементарно включенных мемристоров и диода Зенера с малым энергопотреблением и существенно большей интеграцией элементов по сравнению с активным селектором – транзистором.
2. Разработанные физико-математические модели процессов обработки сигналов и ассоциативного самообучения по правилам долговременной пластичности и пластичности, зависимой от времени импульса, электрическая схема и топология сверхбольшой планарной и 3D

запоминающей матрицы на основе нового компонента наноэлектроники – комбинированного мемристорно-диодного кроссбара.

3. Результаты численного моделирования процессов обработки сигналов при взвешивании импульсов напряжения и сложении токов ячеек в отдельной запоминающей матрице и ассоциативного самообучения запоминающей матрицы в составе аппаратной импульсной нейросети биоморфного нейропроцессора с использованием измеренных характеристик изготовленных мемристоров. Корректность разработанных физико-математических моделей, следующая из согласия результатов численного моделирования с экспериментальными данными и дающая основание для использования этих моделей при проектировании матриц сверхбольшого размера.
4. Генерация новых ассоциаций при численном моделировании нейросети с массивом синапсов, интегрированных в мемристорно-диодный кроссбар запоминающей матрицы, в отличие от ассоциативного самообучения аппаратных нейросетей с дискретными мемристорными синапсами.

Степень достоверности и апробация результатов.

Достоверность разработанных физико-математических моделей и полученных в работе результатов по численному моделированию больших электрических схем, содержащих мемристорно-диодные кроссбары, обеспечена применением известного симулятора SPICE (Simulation Program with Integrated Circuit Emphasis), использованием измеренных характеристик изготовленных мемристоров и согласием расчетных результатов с экспериментальными данными.

Основные результаты работы докладывались и обсуждались на Международных и Всероссийских конференциях в том числе:

- VIII, IX и XII конференциях Нанотехнологического общества России (Москва, 2016, 2017, 2021);

- First International Workshop «Nanoelectronic Memristive Devices for Quantum and Neuromorphic Computing (MEM-Q)» (Kurchatov Institute, Moscow, 2018);
- 6-я Международная научная конференция «Электронная компонентная база и микроэлектронные модули» (Республика Крым, г. Ялта, 2020)
- Международная научная конференция «Применение технологий виртуальной реальности и смежных информационных систем в междисциплинарных задачах FIT-M 2020» (г. Москва, 2020)
- Международная научно-техническая конференция «Smart energy systems 2021» (г. Казань, 2021)

Исследования, выполненные в ходе диссертационной работы, поддержаны грантами РФФИ №19-07-00272 «Электрофизические свойства комбинированного мемристорного-диодного кроссбара - нового компонента наноэлектроники, предназначенного для изготовления запоминающей и логической матриц нейропроцессора» и №19-37-90030 «Генерация нового знания в нейросети на основе массива мемристорных синапсов в запоминающей матрице биоморфного нейропроцессора и принципы увеличения быстродействия и энергоэффективности обработки информации на специализированном устройстве по сравнению с существующими вычислительными средствами».

Публикации. По результатам исследований опубликовано 16 научных работ, в том числе восемь в рецензируемых изданиях: три статьи в журналах, входящих в международные базы; две статьи, входящих в международную базу и в перечень ВАК РФ; одна статья в журнале из международной базы перечня ВАК РФ, а также два патента на изобретение.

Личный вклад соискателя. Разработка физико-математических моделей, весь объем численного моделирования и обработка полученных результатов выполнялись автором лично. Автор выражает благодарность научному руководителю, а также коллегам из НОЦ «Нанотехнологии» за помощь в проведении исследований и обсуждении полученных результатов.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, заключения и списка литературы. Работа изложена на 94 страницах, включает 35 рисунков и 1 таблицу. Список литературы содержит 92 наименований.

ГЛАВА 1. Запоминающие матрицы для хранения информации и обработки входных данных путем взвешивания и суммирования сигналов

В настоящее время в информационных технологиях (ИТ) в области создания искусственного интеллекта разрабатываются искусственные нейросети на простых нейронах и аппаратные средства для традиционных задач распознавания паттернов в видео и аудио информации. Принятие решений в таких устройствах происходит в результате анализа информации, хранящейся в блоках памяти, и выбора наиболее правдоподобного решения на основе заложенных ассоциаций [13]. Включение большого числа нейронов в сеть позволяет увеличить точность распознавания и разрешение входного изображения. Эффективность расчета нейросетей на процессорах общего назначения является низкой. Поэтому для расчета больших нейросетей необходимо использовать суперкомпьютеры или специализированные аппаратные средства.

Другое применение искусственных нейросетей предназначено для моделирования работы отдельных частей мозга. Моделирование биологических нейросетей с большим количеством нейронов требует огромных вычислительных мощностей и сейчас выполняется программно на суперкомпьютерах. Особое внимание уделяется так называемой кортикальной колонке (гиперколонке) универсального элемента неокортекса [14]. Миниколонки предназначены для распознавания элементарных признаков образа и обеспечивает оперативную память, а гиперколонки служат для комплексного объединения элементарных признаков.

Реализации нейросетей можно разделить на три класса: программные, аппаратные и программно-аппаратные (комбинированные). Очевидно, что чем больше доля расчетов выполняется в специализированной электрической схеме программно-аппаратной реализации нейросети, тем больше её быстродействие. Большинство реализаций нейросетей являются

программными и рассчитываются на компьютерах, серверах и суперкомпьютерах.

Примером полностью аппаратной нейросети служит дифракционная глубокая нейронная сеть [15] – многослойный персептрон, в котором невозможна перестройка синаптических связей между нейронами. Обучение таких полностью аппаратных нейросетей производится предварительно перед изготовлением в ходе численного моделирования.

К программно-аппаратному классу относятся нейропроцессоры, работающие как ускорители отдельных математических операций нейросети и построенные как на классической КМОП логике [16, 17] так и с применением мемристорных кроссбаров [18, 19]. В России по проектным нормам 65 нм в 2019 году изготовлен первый отечественный полностью КМОП нейросетевой ускоритель машинного обучения NCore, основанного на математическом аппарате глубоких нейронных сетей для автоматического решения интеллектуальных задач в области распознавания изображений и человеческой речи, управления беспилотными автомобилями и многих других целей [20].

Перечисленные ускорители создаются и используются для ускорения расчетов нейросетей с точечными нейронами. Программно-аппаратные импульсные нейронные сети, например [21] и [22], потребляют меньше энергии, чем традиционные сверточные нейронные сети на точечных нейронах, а также обладают более простой процедурой обучения.

Одним из подходов для увеличения производительности цифровых нейропроцессоров является массовая параллелизация расчетов. В данном случае нейропроцессор представляет собой сеть простых ядер – микропроцессоров, на которых рассчитывается небольшое число нейронов и синапсов. Ускорение вычислений достигается за счет уменьшения влияния узкой шины передачи данных между процессором и памятью из-за наличия в ядрах собственной памяти, в которой хранятся значения синаптических весов. Образцами импульсных нейропроцессоров с такой организацией

являются TrueNorth [22], Brainchip Akida [21, 23] и Intel Loihi [24], изготовленные на транзисторах по КМОП технологии. В последних двух нейропроцессорах уже реализованы механизмы самообучения импульсных нейросетей. Применение в дальнейшем в таких нейропроцессорах мемристоров в качестве синапсов обеспечит значительное сокращение числа транзисторов, благодаря реализации в мемристоре множества резистивных состояний.

Интегрирование мемристивных устройств с КМОП логикой уже применяется при разработке аппаратных нейроморфных сетей [25] и создании чипов памяти [26, 27]. Возможность использования мемристоров в качестве синапсов искусственных нейросетей подтверждена экспериментально [28]. В работе [29] представлен вариант архитектуры нейроморфного сопроцессора на основе мемристорной микросхемы, интегрированной с КМОП логикой. Электрические свойства мемристора, подобные свойствам живого синапса при распространении нервного импульса, описаны в [30]. В этой же работе указана возможность использования мемристора, как суммирующего элемента искусственного нейрона. Использование мемристоров в качестве синаптических связей в подобных устройствах позволит упростить их архитектуру, что увеличит быстродействие и снизит потребление энергии.

Нейропроцессор как вычислительное устройство может быть построен по фон-Неймановской архитектуре с использованием логической [31] и запоминающей [32] матриц на основе мемристоров. Однако, целесообразнее использовать запоминающее устройство, в котором помимо собственно хранения информации обеспечивается ее обработка в виде взвешенного суммирования входных сигналов [33-35], поскольку нейросеть состоит из вычислительных единиц, которые выполняют эту операцию.

Исходя из большой архитектуры нейропроцессора и соответствующего большого количества элементов в электрической схеме, к его узлам предъявляются общие требования: высокая степень интеграции элементов при объединении их в сверхбольшую матрицу; минимизация площади, которую

занимает ячейка матрицы на кристалле; высокие быстродействие и энергоэффективность. Монолитная трехмерная интеграция памяти на мемристорах и логических схем может значительно улучшить интеграцию элементов, производительность и энергоэффективность масштабируемых вычислительных систем и может служить технической основой для создания нейропроцессора. Работоспособность таких схем на мемристорах показана в [36].

Часть расчета нейросети по взвешиванию и суммированию может быть реализована в запоминающей матрице путем умножения напряжения входного сигнала на сопротивление с помощью закона Ома и суммирования получившихся токов по первому закону Кирхгофа. А нейросетевая задача сравнения суммы сигналов с порогом может быть реализована в логической матрице нейропроцессора.

В цифровой запоминающей матрице информация только хранится, а биоморфная матрица позволяет помимо собственно хранения данных в требуемом виде, еще и выполнять операции над входными сигналами. Кодирование передаваемой информации импульсами определенной амплитуды и длительности (подобно потенциалу действия в живом нейроне) позволяет помимо двух крайних состояний мемристора задействовать и промежуточные в качестве синаптического веса. Нейроморфный подход подразумевает обработку информации путем прохождения её через нейронно-синаптическую сеть, в которой происходит взвешивание, суммирование и сравнение суммы сигналов с порогом компаратора. В аппаратной реализации нейропроцессора мемристор может использоваться не только как взвешивающий элемент в запоминающей ячейке, но и как суммирующий элемент в периферийном устройстве матрицы [30].

Существуют энергонезависимые запоминающие устройства на основе ячеек биполярного [26, 37] и униполярного [38] резистивного переключения.

В работе [37] представлен тестовый чип на 64 Мб, построенный по технологии 0,13 мкм с размером ячейки 0,17 мкм². В топологии чипа мемристорный кроссбар расположен над слоем КМОП логики.

Продемонстрированы процедуры записи и чтения. Показано, что предложенная технология СМОх (conductive metal oxide – проводящий оксид металла) совместима электрически и технологически с современной КМОП электроникой.

Тестовая микросхема ReRAM [26] объемом 32 Гб была разработана с использованием 24-нм техпроцесса с диодом в качестве селективного элемента и оксидом металла в качестве материала с резистивным переключением. Массив памяти построен с использованием топологии кроссбар, что позволяет размещать несколько слоев памяти над вспомогательными схемами и минимизировать накладные расходы на площадь чипа. Эффективность кристалла дополнительно повышается за счет совместного использования линий слов и битов соседних блоков. Поскольку количество усилителей считывания под массивом памяти ограничено, применяется конвейерная схема управления массивом, чтобы компенсировать влияние на производительность при использовании быстрого времени переключения ячеек ReRAM. Поскольку в потребляемом кристаллом токе преобладает утечка в матрице, которая определяется напряжением смещения матрицы и условиями эксплуатации, для оптимального энергопотребления введена схема управления каскадом подкачки заряда для динамической адаптации к рабочим условиям. Интеллектуальное считывание и схема компенсации тока утечки во время программирования применяются к архитектуре с использованием больших блоков и обеспечивают плотность элементов микросхемы, которая на несколько порядков выше, чем в предыдущих разработках ReRAM.

В патенте [38] представлено несколько вариантов энергонезависимых цифровых запоминающих устройств с униполярными мемристорами и селективными диодами. Приводится несколько вариантов топологии объединения ячеек в блоки для минимизации площади на чипе.

Устройства [26, 37, 38] используются только для хранения двоичных данных, и архитектура устройства не позволяет выполнять какую-либо обработку входных сигналов.

Использование кроссбара из комплементарных мемристоров в качестве запоминающей матрицы с возможностью вычислений непосредственно в памяти в виде взвешенного суммирования впервые представлено в [32, 39]. В [32] описывается общий дизайн архитектуры кроссбара, основанный на комплементарной ячейке памяти с резистивной коммутацией 2М (рис.1).

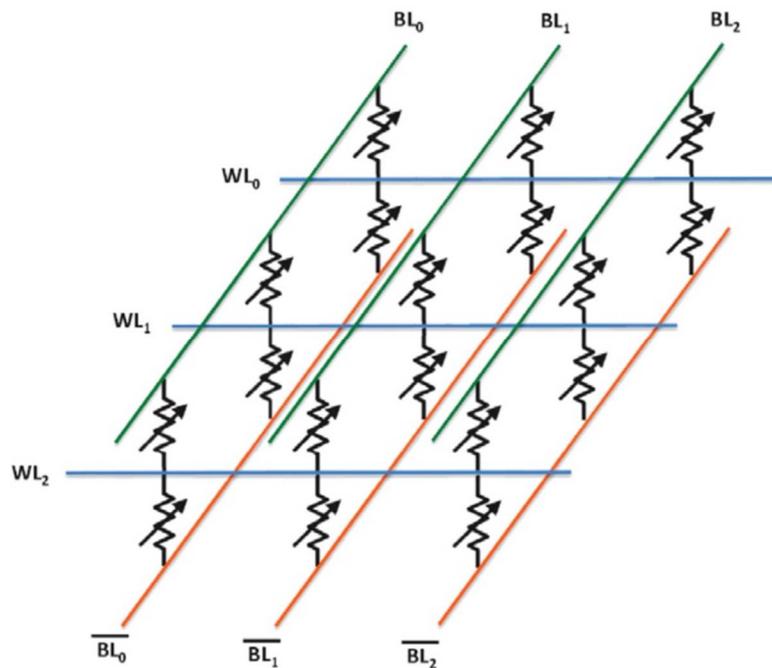


Рис.1. Кроссбар с ячейками 2М [32] для взвешивания входных напряжений в мемристоре и суммирования токов на выходных проводниках.

Эта архитектура позволяет использовать меньшее количество транзисторов и контактов между ячейками памяти и КМОП схемами управления. Комплементарная ячейка и параллельное считывание данных смягчают влияние паразитных токов в массиве. Эта общая архитектура реализована для двух новых технологиях энергонезависимой памяти: STTMRAM (spin-transfer torque magnetoresistive random-access memory – магниторезистивная память с произвольным доступом и записью на основе переноса спинового момента) и OxRRAM (Oxide Resistive Random Access

Memory - память с произвольным доступом на основе оксидов с резистивным переключением). Были разработаны компактные модели резистивных элементов STT-MRAM и OxRRAM для моделирования полной схемы запоминающего устройства. Результаты моделирования STT-MRAM 4×4 и массива OxRRAM 32×32 (длина слова 16 бит) подтверждают функциональность предложенной архитектуры с использованием набора для проектирования КМОП 40 нм. Моделирование отказоустойчивости и анализ производительности этой архитектуры с кроссбаром также были исследованы с учетом вариаций характеристик при изготовлении как ячеек с мемристорами, так и КМОП транзисторов. Предложенная запоминающая матрица может использоваться для реконфигурируемой логики или вычислительных парадигм на основе нейронных сетей.

Работа [39] рассматривает применение этой матрицы в нейронной сети в парадигме RC (Reservoir Computing - резервуарные вычисления) и ELM (Extreme Learning Machine – машина экстремального обучения). RC и ELM предполагают наличие слоя нейронов со случайными значениями синаптических коэффициентов: для RC это входной слой, для ELM – внутренний скрытый слой. При объединении двух стандартных нанoeлектрических структур – кроссбаров со скрытыми нейронами, выполняющими пространственно-временную активацию / интеграцию, при моделировании были продемонстрированы многие функции программных систем RC / ELM: высокая производительность, онлайн-обучение и устойчивость к несовершенствам входных векторов или динамике нейронов.

В матрице реализована последовательная (поочередная) запись информации в комплементарные мемристорные ячейки и параллельное (построчное) считывание их состояния. Комплементарные ячейки уменьшают паразитные токи в кроссбаре при параллельном считывании. При этом в режиме записи необходимо поддерживать на невыбранных ячейках матрицы электрический потенциал, равный половине напряжения записи, что приводит к повышенному потреблению энергии.

В последующих работах [33-35] эта матрица, позволяющая взвешивать и суммировать входные сигналы в виде постоянных напряжений, применена в качестве аппаратной реализации персептрона. Она же использовалась в [39], где на основе биоморфного подхода выполнялось суммирование входных импульсов напряжения.

В [33] изучена устойчивость этой архитектуры к большинству типов дефектов составляющих устройств и изменчивости параметров при изготовлении с помощью моделирования методом Монте-Карло. Наблюдается хорошая устойчивость к неидеальности устройств, даже если на первый взгляд различия критические. Добавив разумную избыточность и используя технику «конкурентного обучения», можно избежать этой проблемы и достичь высокой степени устойчивости для небольшой системы. Кроме того, также разработаны вероятностные модели для понимания и прогнозирования влияния дефектов и изменчивости устройства в очень крупномасштабных нейрочипах. Результаты очень выгодно отличаются от предыдущих работ по избыточному подходу. Исследование не опирается на конкретную технологию, поэтому авторы не могут дать никаких прогнозов производительности с точки зрения масштабируемости, энергии и скорости. Кроме того, рассматривалась только наиболее вероятные статическая изменчивость и дефекты, но эти технологии также будут страдать от изнашивания и временных ошибок.

В статье [34] представлен NC (neural crossbar – нейронный кроссбар), как называют его авторы, на основе мемристора, который реализует обучение с учителем на чипе. Вместо использования стандартного КМОП нейрона, функцию активации реализована на простом КМОП инверторе. Предложена компактная обучающая ячейка, которая состоит из триггера-защелки на основе кроссбара из двух антипараллельно ориентированных бинарных мемристоров. Эта конструкция обеспечивает интеграцию с более высокой плотностью и может быть естественным образом расширена до многослойной нейронной сети. Используя комплект для проектирования

КМОП 40 нм и основанную на физических принципах компактную модель высокопроизводительного сегнетоэлектрического туннельного мемристора, выполнено моделирование переходных процессов, чтобы проверить работоспособность предложенного нейронного кроссбара. Построена многослойная нейросеть путем каскадирования однослойных сетей, что позволяет сети обучаться для воспроизведения нелинейно разделимых функций (например, функции исключающего ИЛИ). Отказоустойчивость оценена также с помощью моделирования методом Монте-Карло.

Показано [35], что рассеивание энергии в предлагаемой схеме в основном зависит от технологии мемристивных наноустройств (диапазон проводимости, характеристики чтения / записи). Большая часть общей рассеиваемой мощности связана с этапом обучения, а часть рассеиваемой мощности, потребляемой во время работы обученной нейросети на порядок меньше. Потреблением в режиме ожидания можно пренебречь благодаря энергонезависимости мемристивных наноустройств.

Однако, разработанная в [32] матрица не может быть использована в качестве сверхбольшой запоминающей матрицы нейропроцессора из-за низкой энергоэффективности мемристорной ячейки при записи и высокой деградации выходного сигнала при считывании. Повышенное потребление энергии возникает из-за паразитных токов через соседние ячейки в кроссбаре. Применение обычного диода в качестве селективного элемента для исключения взаимовлияния ячеек возможно только для униполярных мемристоров, поскольку для перевода биполярного мемристора в непроводящее состояние требуется протекание тока в обратном направлении.

Массивы мемристорных ячеек типа 1Т1М [40-44] с селективным транзистором используются при построении аппаратных нейросетей [3, 45]. Мемристорные кроссбары 1Т1М (рис.2) обладают реконфигурируемыми энергонезависимыми состояниями сопротивления и могут устранить узкие места в скорости и энергоэффективности при векторно-матричном умножении, основной вычислительной задаче при обработке сигналов и

изображений. Однако использование таких систем для умножения вектора аналоговых амплитуд напряжений на матрицу аналоговой проводимости в достаточно большом масштабе оказалось сложной задачей из-за трудностей в разработке устройств и интеграции массивов.

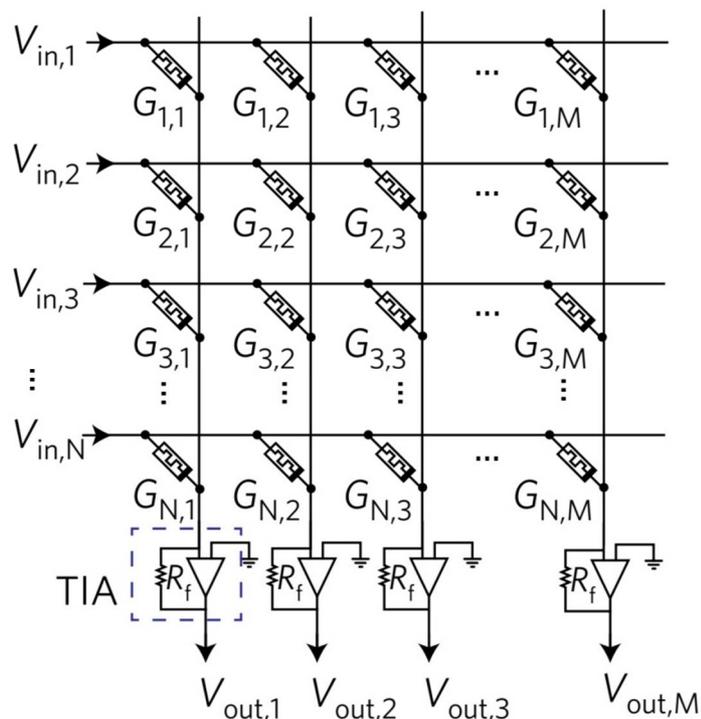


Рис.2. Кроссбар ячеек 1Т1М [40] для взвешивания входных напряжений в мемристоре и суммирования токов на выходных проводниках. Для простоты селекторные транзисторы не показаны на рисунке.

В [40] показано, что реконфигурируемые мемристормые кроссбары, состоящие из мемристоров из оксида гафния поверх МОП транзисторов, способны к аналоговому умножению векторной матрицы с размерами массива до 128×64 ячеек. Демонстрируемая точность компонент выходного вектора (5–8 бит, в зависимости от размера массива) является результатом высокой доли пригодных устройств в массиве (99,8%) и многоуровневого стабильного состояния мемристоров при линейных вольт-амперных характеристиках и низком сопротивлении проводников между ячейками. С помощью больших мемристормых кроссбаров продемонстрирована обработка сигналов, сжатие изображений и сверточная фильтрация.

Энергоэффективность системы составила более 119,7 триллиона эквивалентных операций в секунду на ватт при времени считывания 10 нс, и ожидается, что она значительно возрастет с увеличением размерности векторов и матриц и с улучшением схемотехники.

Для аппаратной реализации импульсной нейросети предложена ячейка типа 2Т1М [4], в которой второй транзистор используется в механизме самообучения. Входной сигнал для ячейки сложный и состоит из двух импульсов: импульс напряжения на один из контактов мемристора и импульс на затвор одного из транзисторов (рис.3).

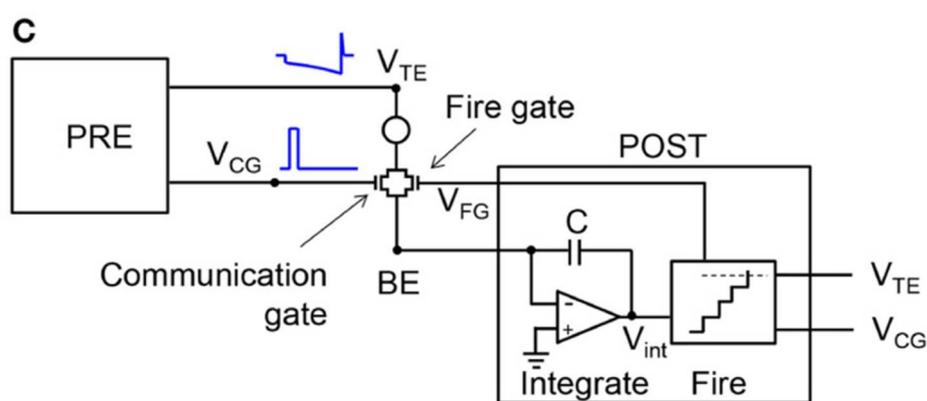


Рис.3. Использование ячейки 2Т1М [4] в качестве синапса импульсной нейросети.

При срабатывании нейрона он посылает на второй затвор ячейки импульс напряжения для подстройки веса. Импульсы по форме практически повторяют биологические. В работе [46] предложена ячейка типа 1Т1М с новым типом полевого транзистора, который обладает малым размером и может изменять тип носителей заряда в проводящем канале, что позволяет свести 2Т1М ячейки на обычных транзисторах к ячейке типа 1Т1М. В этой работе исследованы возможности применения транзисторов с управляемой полярностью (PCT – Polarity Change Transistors), для улучшения работы биполярных массивов памяти RRAM энергонезависимой памяти. Показано, что стандартные массивы памяти на основе КМОП требуют высокого напряжения на затворе (> 2 В) или огромной битовой ячейки ($3 \times$) для выполнения субмикросекундной операции

сброса. В этом контексте предложено две инновационные битовые ячейки типа 1T1R, использующие PCT и обеспечивающие быстрый сброс при обеспечении плотной организации (область битовых ячеек от $16,4F^2$ до $25,76F^2$). Предлагаемые битовые ячейки промоделированы с помощью компактных моделей SiNWFET PCT и OxRAM RRAM и сравниваются с технологией 28 нм КМОП КНИ (кремния на изоляторе) с низким энергопотреблением. Предлагаемые битовые ячейки превосходят характеристики битовых ячеек 1T1R и 2T1R при сбросе (улучшение скорости от 5 до 10^5 раз), занимают меньшую площадь (уменьшение площади от 1,35 до 2,6 по сравнению с 2T1R) и позволяют избежать перегрузки затвора (1,2 В против более чем 2 В в битовых ячейках 1T1R), что снижает проблемы с надежностью селектора. Также, предлагаемые битовые ячейки работают на 15–67% быстрее, чем битовые ячейки КМОП при чтении.

Такие ячейки [2, 42-44, 46] обладают сравнительно низкой интеграцией из-за наличия дополнительных управляющих проводников и большой площади транзисторов по сравнению с размерами мемристоров.

В статье [47] предложена ячейка 4M1M (рис.4), содержащая селектор на основе мемристоров, который может использоваться в цифровых периферийных схемах чтения / записи памяти без возможности взвешенного суммирования. Кроме того, обсуждаются методы последовательного чтения с длинными интервалами обновления и быстрой записи для предлагаемой конструкции.

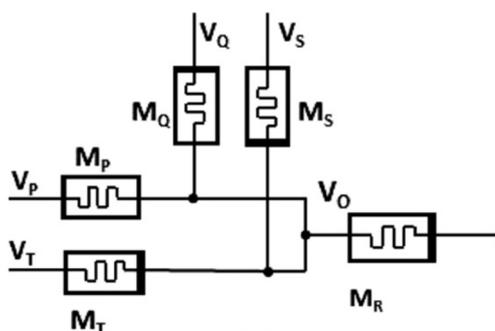


Рис.4. Электрическая схема ячейки 4M1M [47] с двумя селекторами, каждый из которых образован двумя мемристорами.

Предложенная архитектура имеет ряд отличных характеристик, таких как высокая плотность, энергонезависимость, низкое энергопотребление и хорошая масштабируемость. Результаты моделирования также показывают, что предложенная матрица памяти имеет лучшие характеристики по сравнению с другими матрицами на основе мемристоров, предложенными в существующей технической литературе. Ячейка содержит два селектора из комплементарных пар мемристоров. Ток утечки в массиве [47] такой же, как для 1Т1М ячейки.

Архитектура матрицы 2М1М [48], способная работать как цифровая память и логическая схема, основана на ячейке без транзисторов. Предлагаемая ячейка памяти состоит из одного целевого мемристора и двух мемристоров доступа (рис.5), которые образуют селектор для уменьшения влияния паразитных токов.

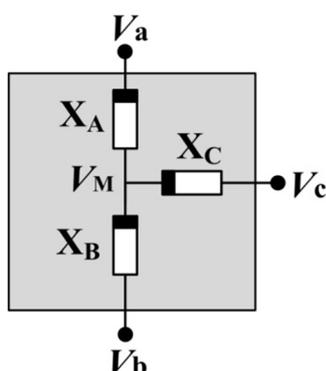


Рис.5. Электрическая схема ячейки 2М1М [48], с двумя селекторами, каждый из которых образован одним мемристором.

Эта структура имеет значительно большую плотность разводки и меньшее количество мемристоров на бит по сравнению с аналогами. Следовательно, это может быть подходящая структура для памяти с высокой плотностью памяти и логических приложений. В дополнение к возможностям вычисления в памяти структура 2М1М обеспечивает более высокую плотность и меньшее энергопотребление по сравнению с традиционной статической памятью с произвольным доступом на основе КМОП. По сравнению с предыдущими работами, результаты моделирования показывают значительное

улучшение реализации ячейки памяти с точки зрения времени записи (1,11 нс), времени чтения (200 пс), плотности ($80 \text{ Гб} / \text{см}^2$), энергопотребления ($23,2 \times 10^{-3} \text{ фДж} / \text{бит}$), и сложности разводки. Кроме того, паразитный ток равен 90 нА при операции чтения из ячейки памяти, что значительно ниже по сравнению с аналогами. Однако, при моделировании не учитывались паразитные реактивные элементы в кроссбаре (емкости и индуктивности). Другой недостаток работы заключается в том, что мемристор состояния и селективные мемристоры должны обладать разными электрическими характеристиками: напряжением переключения и предельными сопротивлениями. Использование разных по свойствам мемристоров усложняет изготовление такого массива.

Кроме этого, при каждом считывании и записи в [47, 48] нужно изменять состояние селективных мемристоров, что негативно сказывается на надежности из-за быстрого износа. Состоящие только из мемристоров ячейки 4М1М [47] и 2М1М [48] позволяют производить логические операции и хранить двоичные данные, но не предназначены для использования в аналоговых вычислениях.

С точки зрения масштабирования сверхбольшой запоминающей матрицы необходим выбор селективного элемента малого размера. Ячейки типа 1S1М [49-52] и 1D1М [53-55] являются наиболее оптимальными с точки зрения занимаемой площади и нелинейности ВАХ [49].

Ячейка [49] состоит из вертикально объединенных МІМ (Metal-Insulator-Metal – металл-диэлектрик-металл) селектора и биполярного мемристора на основе диоксида титана (рис.6), её площадь составляет $4F^2$.

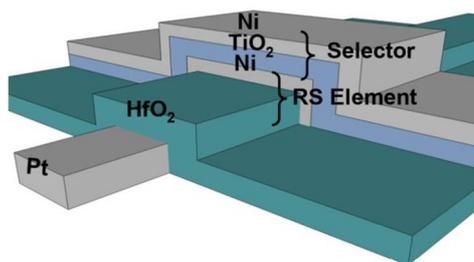


Рис.6. Структура ячейки 1S1М [49], содержащая селектор (Ni / TiO₂ / Ni) с симметричной вольт-амперной характеристикой.

Экспериментально изготовлен кроссбар 8x8 на гибкой подложке. На основе экспериментально измеренных характеристик малого кроссбара и последующего моделирования показано, что деградация напряжения входного сигнала достигает уровня 30% наблюдается при 1000x1000 ячеек.

В [50] представлен двунаправленный селектор TiN / аморфный кремний / TiN (MSM – Metal-Semiconductor-Metal) на основе туннельного эффекта через барьеры Шоттки для массивов RRAM высокой плотности. Устройства демонстрируют превосходные характеристики с высоким током возбуждения, превышающим $1 \text{ mA} / \text{cm}^2$, и высокой нелинейностью. Надежность продемонстрирована на структурах размером 40 нм, со статистической способностью выдерживать более 10^6 циклов биполярного переключения мемристоров. Стабильность работы устройства составляет более 3 часов при 125°C . Экспериментально изучались только селекторы. Моделирование больших кроссбаров показывает деградацию напряжения до 30% при 10^5 ячеек при напряжении полувыбранных ячеек в 1/2 от напряжения выбранной ячейки.

Селективные элементы, предложенные в [49, 50] обладают симметричной вольт-амперной характеристикой, что не позволяет их использовать для реализации логических операций. Соответственно, в запоминающей матрице также не желательно их использовать поскольку это приведет к дополнительным шагам при производстве нейропроцессора из-за использования разных материалов.

Существуют структуры материалов для мемристоров с сильно несимметричной вольт-амперной характеристикой. Такие структуры можно рассматривать как вариант 1S1M. Селектор в данном случае имеет общие слои со структурой мемристора. В [56] свойство самовыпрямления запоминающих устройств возникает из-за асимметричных барьеров для туннелирования электронов в ловушки слоя Si_3N_4 , который зажат между верхним и нижним электродами, имеющими разную работу выхода. В работе [57] разработано КМОП-совместимое устройство RRAM не требующее

формовки с самовыпрямлением и структурой Pd / HfO₂ / WO_x / W. Выпрямляющие характеристики были получены благодаря контакту Шоттки Pd / HfO₂ и квазиомическому контакту W / WO_x. Оценок влияния полученной нелинейности на размер кроссбара в обеих работах не приводится.

Участок с высокой нелинейностью ВАХ ячейки необходим для исключения паразитных токов при записи в больших мемристорных кроссбарах. Для этого можно использовать отдельный селективный элемент или мемристоры с самовыпрямляющим свойством [56, 57]. Поскольку в ячейке запоминающей матрицы [32] пара комплементарных мемристоров образуют резистивный делитель напряжения, то для обеспечения линейной зависимости выходного напряжения делителя от входного напряжения нелинейный селектор должен быть отдельным элементом. Самовыпрямляющие мемристоры можно было бы применить в логической матрице 1D1M [1], но при производстве целесообразнее использовать унифицированные элементы для уменьшения количества технологических операций при производстве нейропроцессора [1], который содержит обе матрицы.

Примером создания 3D архитектуры запоминающей матрицы с использованием мемристоров и диодов является энергонезависимое устройство на основе ячеек униполярного резистивного переключения, включенных последовательно с нелинейным элементом [38]. В такой схеме исключение взаимовлияния соседних ячеек реализовано с помощью обычного полупроводникового диода. Предложенная в патенте топология предполагает большое число производственных операций из-за горизонтального относительно подложки расположения областей с разной технологией изготовления.

Принципы ассоциативного самообучения и формирование новой ассоциации (нового знания) в импульсной нейросети с мемристорными синапсами по правилу Хебба впервые представлены в [31]. Основная идея заключается в использовании двунаправленных нейронов, которые посылают

сигнал о своей активации обратно в синапсы. Аппаратные реализации этой идеи в виде нейросети из трех нейронов с синапсами на дискретных мемристорах предложены в последующих работах [4, 9, 58-64]. При экспериментальной демонстрации ассоциативного самообучения в нейросети с мемристорами на основе органического полимера – парилена [62] явно показана генерация новых импульсов, связанных с возникновением новой ассоциации. В работе [63] проводилось численное моделирование нейросети с мемристорными синапсами [62], характеристики которых были предварительно получены экспериментально.

В [4] Входной сигнал для ячейки сложный и состоит из двух импульсов: импульс напряжения на один из контактов мемристора и импульс на затвор одного из транзисторов. При срабатывании нейрона он посылает на второй затвор ячейки импульс напряжения для подстройки веса. Импульсы по форме практически повторяют биологические. Реализовано обучение по правилу синаптической пластичности, зависимой от времени импульсов, и приведена соответствующая кривая пластичности. Собрана аппаратная нейросеть с 4x4 входами и одним выходом.

Прототип аппаратного двухслойного персептрона с синапсами на основе дискретных комплементарных мемристоров [10] изготовлен в Нижегородском университете им. Лобачевского в рамках проекта РФФ №16-19-00144 и предназначен для задач нелинейной классификации при обучении с учителем. В Курчатовском институте исследования по созданию нейропроцессоров представлены реализацией однослойного и двухслойного аппаратных персептронов на базе полианилиновых мемристоров [65, 66].

Однако предложенные электрические цепи аппаратной реализации ассоциативной памяти не могут быть использованы для построения большой аппаратной нейросети с высокими интеграцией элементов и энергоэффективностью. Причиной является отсутствие интеграции мемристоров в кроссбары и наличие в схемах нейронов и синапсов большого числа активных электронных элементов с высоким энергопотреблением.

Существуют аппаратные реализации нейросетей с мемристорными массивами в виде кроссбаров [3, 5-7]. Онлайн-обучение на нейроморфной сети на основе RRAM было продемонстрировано экспериментально [45] с использованием изготовленного 1к-битного массива 1T1R (128x8 ячеек) и архитектуры однослойного персептрона. Нейросеть показывает низкое энергопотребление (2,23 нДж) на итерацию и высокую скорость распознавания 89,1% на тестовой выборке с шумами.

В [3] продемонстрирована способность к самоадаптивному обучению *in situ* (лат. «на месте») многослойной нейронной сети, построенной путем монолитной интеграции массивов мемристоров на КМОП подложку. Транзисторы обеспечивают надежное, линейное и симметричное обновление синаптического веса, позволяя обучать сеть с помощью стандартных алгоритмов машинного обучения. После обучения с помощью алгоритма стохастического градиентного спуска на 80 000 изображений, взятых из обучающего набора MNIST, достигнута точность 91,71% на всем тестовом наборе из 10 000 изображений. Эта точность на 2,4% ниже, чем при идеализированном моделировании, несмотря на 11% дефектных мемристоров. Продемонстрированная производительность с онлайн-обучением и выводом на месте позволяет предположить, что мемристорные переключатели являются многообещающей высокоскоростной и энергоэффективной технологией для искусственного интеллекта.

Пятислойная сверточная нейронная сеть (CNN – convolutional neural network) [5] на основе мемристоров для распознавания изображений из датасета MNIST10 достигает высокой точности более 96 процентов. Помимо параллельных сверток с использованием разных ядер с общими входами, было продемонстрировано дублирование нескольких идентичных ядер в массивах мемристоров для параллельной обработки разных входных данных. Эта нейроморфная система на основе мемристоров имеет энергоэффективность на два порядка выше, чем у современных графических

процессоров, и, как показано, может масштабироваться до более крупных сетей, таких как остаточные нейронные сети.

В [6] разработан и произведен полностью функциональный программируемый нейроморфный вычислительный чип, в котором матрица пассивных мемристоров интегрирована с полным набором аналоговых и цифровых компонентов и встроенным процессором. Интегрированный чип позволяет отображать различные нейроморфные алгоритмы и алгоритмы машинного обучения на чипе посредством простого изменения программного кода. Были продемонстрированы три различные и часто используемые модели: персептрон, разреженное кодирование и РСА (principal component analysis – метод главных компонент) с интегрированным слоем классификации. Точность классификации 100% была достигнута для зашумленных греческих букв 5×5 для однослойного персептрона, надежный анализ разреженного кодирования был получен из исчерпывающего набора тестов с использованием шаблонов 4×4 , а уровень классификации 94,6% был экспериментально получен для набора данных скрининга рака груди.

Экспериментальная демонстрация обучения с подкреплением в трехслойной нейросети с запоминающей матрицей 1T1R с использованием модифицированного алгоритма обучения, адаптированного для гибридной аналого-цифровой платформы представлена в [7]. Нейросеть успешно изучила стратегии для различных классических сред управления (тележка-шест и подъем в горку) с минимальными предварительными знаниями, демонстрируя возможность представления синаптических весовых коэффициентов с точностью от 4 до 5 бит с двухимпульсной схемой программирования и записью без проверки. Показано, что по сравнению с обычными цифровыми системами в реальных задачах обучения с подкреплением такой подход обладает потенциалом для повышения производительности и энергоэффективности почти на порядок.

В работах [5-7] нейросети не работают с импульсами, а мемристормый кроссбар используется исключительно в качестве сопроцессора-ускорителя

матричного умножения. Авторы [5-7] используют термин «обучение in situ», что можно трактовать как самообучение, однако этот термин здесь относится к устройству в целом. То есть способность к обучению не заложена непосредственно в систему массив синапсов – массив нейронов, а реализуется отдельной периферийной электрической схемой на основе цифрового микроконтроллера, являющейся составной частью всего устройства. Подобный подход может обеспечить гибкость в реализации процедур обучения, поскольку микроконтроллер является программируемым. Однако, из-за такой организации устройства процедура обучения является отложенной и производится после прохождения информационных сигналов через аппаратную нейросеть. При этом во время выполнения обучения аппаратная нейросеть игнорирует возможные входные сигналы, что может быть нежелательно для некоторых применений, например, в системах управления. Кроме этого в работе [5] реализовано только обучение с учителем.

Биоморфное самообучение предполагает подстройку весов непосредственно во время прохождения информации по нейросети без периодов невосприимчивости к входной информации. В работе [7] высказывается мысль о возможной реализации непрерывного самообучения в нейросети с кроссбаром при использовании двунаправленных нейронов, которые посылают сигнал о своей активации обратно в синапс, но не приводится информации о возможных реализации такого обучения.

1.1. Выводы к главе 1.

Большинство реализаций нейросетей являются программными и рассчитываются на компьютерах, серверах и суперкомпьютерах. Существующие программно-аппаратные нейропроцессоры работают как ускорители отдельных математических операций нейросети в неимпульсных нейросетях и построены как на классической КМОП логике, так и с применением мемристорных кроссбаров. Программно-аппаратные импульсные нейронные сети потребляют меньше энергии, чем традиционные

сверточные нейронные сети на точечных нейронах, а также обладают более простой процедурой обучения.

Существующие аппаратные средства в виде запоминающих матриц с мемристорным кроссбаром выполняют узкоспециальные функции, в основном матричное умножение, которое представляет собой взвешенное суммирование входных импульсов напряжения. Возможность обучения нейросети в таких устройствах не предусмотрена. Кроме того, рассмотренные запоминающие матрицы обладают низкой интеграцией элементов. Активные селекторы – транзисторы увеличивают энергопотребление и требуют дополнительных управляющих проводников. Применение мемристоров в качестве селекторов подразумевает изменение их состояния при каждом чтении и записи в ячейку, что приводит к быстрому износу. Селективные элементы на основе барьеров Шоттки обладают симметричной вольт-амперной характеристикой, что не позволяет их использовать для реализации логических операций. Соответственно, в запоминающей матрице также не желательно их использовать поскольку это приведет к дополнительным шагам при производстве нейропроцессора из-за использования разных материалов.

Процедура обучения нейросети в аппаратных средствах, содержащих в себе интегрированные мемристорные кроссбары, реализована в [5-7] и является отложенной из-за программного расчета подстройки весов в периферийной системе после прохождения входных сигналов, что приводит к потере информации. Возможность обучения нейросети как с учителем, так и по правилу Хебба, в этих системах реализовано внешней программируемой системой. Аппаратное ассоциативное самообучение без потерь входной информации и формирование новой ассоциации в импульсной нейросети реализовано только с использованием дискретных мемристорных синапсов [8-12]. Таким образом, в настоящее время не существует аппаратных реализаций импульсных нейросетей на основе массивов мемристоров, интегрированных в мемристорные кроссбары, с биоморфным самообучением.

ГЛАВА 2. Запоминающая матрица для аппаратной нейросети биоморфного нейропроцессора.

2.1. Физические основы создания ячейки сверхбольшой запоминающей матрицы

Исходя из большой архитектуры нейропроцессора и соответствующего большого количества элементов в электрической схеме, к его узлам предъявляются общие требования: высокая степень интеграции элементов при объединении их в сверхбольшую матрицу; минимизация площади, которую занимает ячейка матрицы на кристалле; высокие быстродействие и энергоэффективность. Монолитная трехмерная интеграция памяти на мемристорах и логических схем может значительно улучшить интеграцию элементов, производительность и энергоэффективность масштабируемых вычислительных систем и может служить технической основой для создания нейропроцессора. Работоспособность таких схем на мемристорах показана в [67].

Комплементарное включение двух мемристоров позволяет постоянно поддерживать высокое входное сопротивление ячейки в рабочем режиме, что обеспечивает ее низкое энергопотребление. Поскольку комплементарные мемристоры пространственно находятся в одном активном слое рядом друг с другом, то влияние неоднородности их характеристик на выходное напряжение ячейки будет минимальным.

При использовании комплементарных мемристоров изменение их состояния в ячейке производится последовательно. При этом возможны три комбинации высокоомного и низкоомного состояний мемристоров в паре. Таким образом, в ячейку можно записать три состояния, если мемристоры работают в режиме ключа и больше трёх состояний, если мемристорный материал обеспечивает плавное переключение и несколько устойчивых значений проводимости.

Так как предполагается работа в импульсном режиме, то при объединении таких ячеек в матрицу необходимо решить проблему

возникновения паразитных токов. Применение обычного диода для исключения взаимовлияния ячеек возможно только для униполярных мемристоров, поскольку для перевода биполярного мемристора в непроводящее состояние требуется протекание тока в обратном направлении. Использование диода Зенера, имеющего низкое напряжение обратимого пробоя, устраняет эту проблему.

Основные требования к мемристору – высокое отношение предельных электрических сопротивлений, позволяющее реализовать множество резистивных (синаптических) состояний, и низкие напряжения переключения, обеспечивающее малое энергопотребление схемы. Требуемыми характеристиками диода Зенера являются низкие напряжения открытия в прямом смещении и при пробое, высокое отношение электрических сопротивлений в закрытом и открытом состояниях, а также при обратимом пробое (сильная нелинейность).

2.2. Электрическая схема ячейки на основе комплементарных мемристоров.

Ячейка кроссбара (рис.7) содержит два мемристора M1 и M2 с общим электродом, совмещенным с анодом диода Зенера D1 [68, 69].

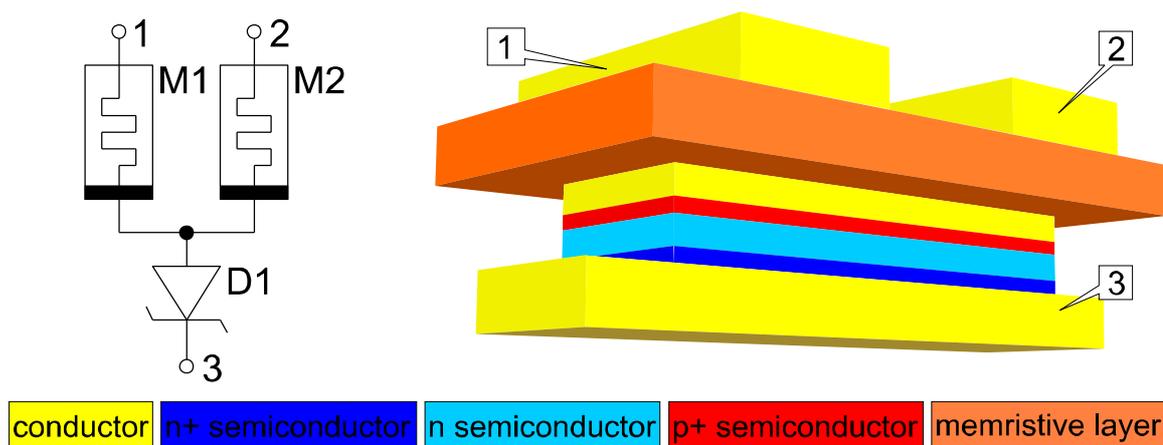


Рис.7. Электрическая схема и топология отдельной ячейки кроссбара.

Применение диода Зенера в качестве селективного элемента, подключенного электрически последовательно к мемристор, позволяет

увеличить размер кроссбара за счет уменьшения паразитных токов между ячейками, а также уменьшить деградацию выходного сигнала при суммировании входных импульсов напряжения в матрице. Выбор в пользу пассивного элемента – диода по сравнению с традиционно применяемым в качестве селективного активного элемента – транзистора сделан из-за его меньших размеров и энергопотребления. Кроме этого транзистор требует наличия дополнительного управляющего проводника.

На нижнем электроде ячейки 3 расположены диоды Зенера, состоящие из последовательно наносимых слоев сильнолегированного n-полупроводника, слаболегированного n-полупроводника и сильнолегированного p-полупроводника. На аноде диода располагается общий металлический электрод комплементарных мемристоров, непосредственно соприкасающийся с вышележащим сплошным мемристорным слоем, на котором расположены верхние электроды ячейки 1 и 2. Комплементарное включение двух мемристоров позволяет постоянно поддерживать высокое входное сопротивление ячейки в рабочем режиме, что обеспечивает ее низкое энергопотребление.

Мемристорный слой и полупроводниковые слои диода могут быть изготовлены в магнетронном технологическом модуле. Слои полупроводников с донорной или акцепторной примесью и разным уровнем легирования создаются путем одновременного распыления катодов из материалов чистого полупроводника и легирующей примеси [70, 71].

2.3. Электрические схемы планарной двухслойной и 3D матрицы на основе интеграции мемристорных ячеек.

На основе двух топологий комплементарной мемристорно-диодной ячейки разработаны два варианта запоминающей матрицы – с параллельным и последовательным выводом данных (рис.8) [68, 72].

Линии X и Y на рис.8б, накрест пронизывая объем матрицы второго типа, объединяют ячейки в электрическую сеть по принципу построения

кроссбаров. Каждая ячейка включена в перекрестье для организации побитного доступа, по аналогии с традиционными схемами DRAM, принципы функционирования которых описаны в технической литературе [73].

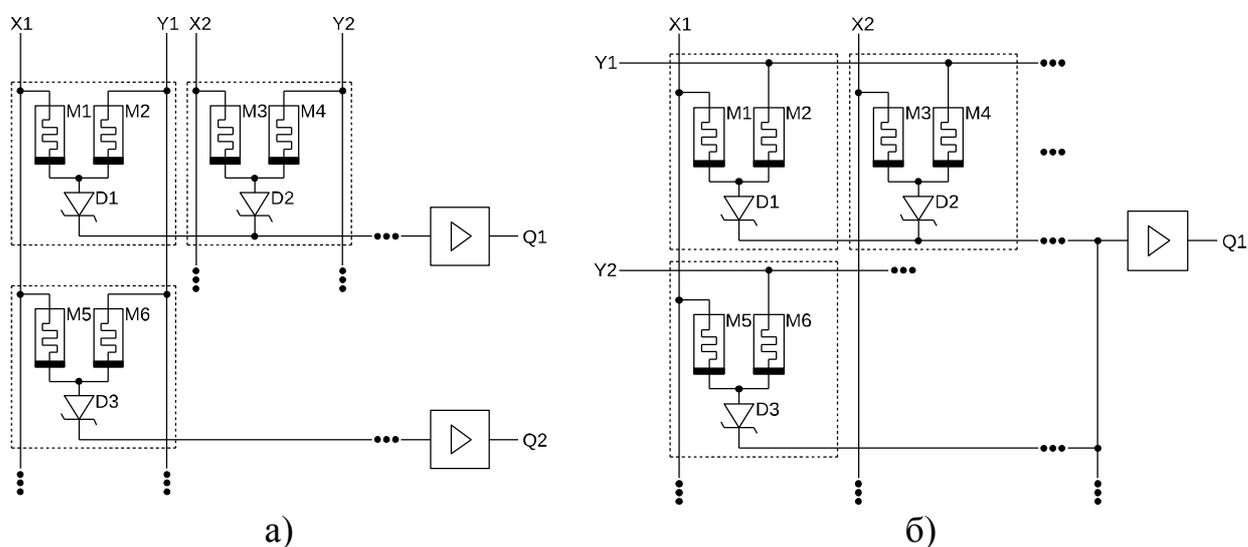


Рис.8. Матрица запоминающего устройства с параллельным а) и с последовательным выводом данных б) через общую шину.

Информация из матрицы считывается последовательно при помощи входного драйвера строк по нижней шине Q, на которую сигнал подается через диод выбранной ячейки, при этом диоды остальных ячеек остаются в закрытом состоянии.

Входные драйверы представляют собой усилители сигналов с мемристоров и формирователи уровней напряжений для дальнейшей их передачи в последующие логические устройства. Выходные КМОП драйверы собраны по классической схеме, основные принципы реализации которых представлены в работе [74]. Драйверы выполняют функции подачи на шины питания: высокого надпорогового напряжения для записи верхних или нижних мемристоров и низкого подпорогового напряжения для считывания данных через объединенные катоды диодов с помощью входных драйверов.

Логика работы выходных драйверов заключается в последовательной подаче импульса тока для закрытия открытого мемристора, а затем импульса напряжения для открытия другого мемристора. При этом общее

сопротивление пары всё время удерживается высоким, а сквозной ток через комплементарные мемристоры остается минимальным, что повышает энергоэффективность всей матрицы. Подача импульса тока осуществляется через прямосмещённый диод Зенера, а импульс напряжения подается при лавинном пробое на обратной полярности. Первая формовка ячеек, необходимая для установления рабочей полярности многих мемристивных материалов, выполняется аналогично рабочим переключениям с помощью выходных драйверов, но на большей длительности и амплитуде импульсов.

Отличающийся более высокой скоростью принцип переключения комплементарных ячеек без участия диода Зенера возможен только для матрицы с побитным доступом, показанной на рис.8б. В этом случае, с помощью драйверов управляющих сигналов подаётся импульс надпорогового напряжения на выбранную в перекрестье ячейку, при этом в комплементарной паре закрытый мемристор открывается, а затем другой мемристор через первый закрывается. Не исключена обратная последовательность переключения комплементарных мемристоров.

Одна и та же электрическая схема и некоторые отличия топологии двух разработанных типов комплементарных мемристорно-диодных ячеек допускают создание запоминающей матрицы с параллельным и последовательным доступом к записи и считыванию данных. Предложенные топологии мемристорно-диодных ячеек дают возможность добиться высокой степени интеграции при объединении их в сверхбольшую матрицу, в которой крупные КМОП-транзисторы являются общими для больших строк ячеек. При этом вся площадь матрицы заполняется мемристорными ячейками нанометрического размера, а крупные элементы вынесены на периферию и не расходуют площадь кристалла. Разработанная конструкция матрицы решает проблему взаимного влияния узлов, характерную для мемристорных схем кроссбаров, так как общее сопротивление ячеек, основанных на комплементарных мемристорах, всегда остаётся высоким. При этом через

ячейки протекает минимальный ток, что и определяет энергоэффективность матрицы.

Трёхмерная запоминающая матрица представляет собой объединение планарных двухслойных запоминающих матриц в 3D структуру из одинаковых горизонтально расположенных и зеркально ориентированных по отношению друг к другу комбинированных кроссбаров, причем шины строк или столбцов соседних кроссбаров являются общими [75]. Двухслойный комбинированный кроссбар можно рассматривать как отдельный функциональный пласт. На рис.9 приведена электрическая схема фрагмента трехмерной запоминающей матрицы из трех комбинированных кроссбаров, поясняющая принцип соединения соседних пластов.

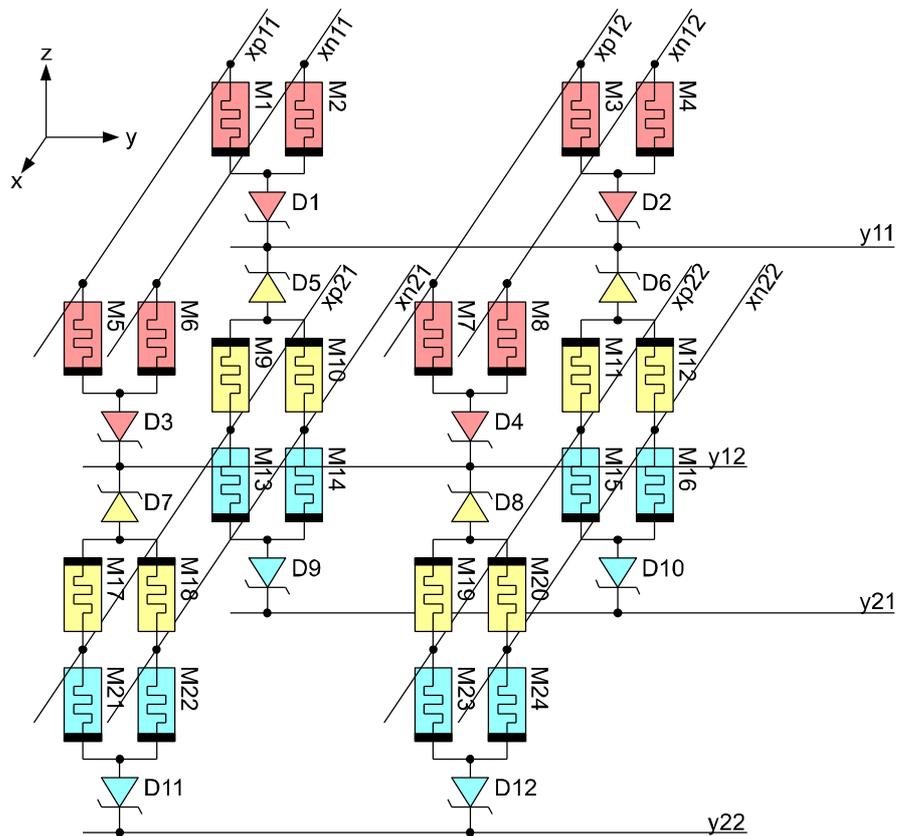


Рис.9. Электрическая схема фрагмента трехмерной запоминающей матрицы.

Три цвета соответствуют трём комбинированным кроссбарам.

Предлагаемая 3D запоминающая матрица на основе комбинированного мемристорно-диодного кроссбара работает в двух режимах: режиме записи и режиме чтения данных.

При записи данных в любую ячейку последовательно во времени на каждый мемристор для изменения его проводимости подается напряжение, большее порога переключения мемристора. Например, для ячейки, состоящей из комплементарных мемристоров M1 и M2, это напряжение поступает с одной стороны через проводящие шины xp11 и xn11, являющимися контактами мемристоров. С другой стороны напряжение подается через диод Зенера D1 по выходной шине y21 с программирующего драйвера, расположенного на периферии матрицы. Диод Зенера выполняет функцию управляемого напряжением переключателя и работает в режиме пробоя во время записи мемристора. В такой схеме запись данных может производиться одновременно в несколько ячеек по шинам, расположенным параллельно.

В режиме чтения матрицы на контакты 1, 2 выбранной ячейки относительно общей точки схемы подаются низкие напряжения противоположной полярности, абсолютная величина которых меньше порогового напряжения переключения мемристора. В результате комплементарная пара мемристоров образует резистивный делитель напряжения. Напряжение со средней точки делителя через диод Зенера поступает на периферийное суммирующее устройство.

2.4. Топология комбинированного мемристорно-диодного кроссбара.

Топология комплементарной мемристорно-диодной ячейки, позволяющей организовать матрицу запоминающего устройства с параллельным выводом данных, представлена на рис.10 [71, 72]. На подложке в изолирующем материале создана матрица вертикально ориентированных диодов с катодами внизу, которые объединены построчно проводниками, показанными в плоскости чертежа топологии горизонтальными линиями. Методом вакуумного магнетронного осаждения сверху на аноды диодов наносится слой мемристивного материала на основе оксида титана. При этом анод диода, являющийся областью n-типа,

располагается под мемристивным слоем и представляет собой общий контакт комплементарных мемристоров в соответствии с электрической схемой (рис.10а). Комплементарные мемристоров образованы внутри активного слоя между анодным контактом диода и двумя верхними проводниками, которые сверху показаны уходящими линиями на топологическом чертеже (рис.10б).

Второй тип топологии комплементарной мемристивно-диодной ячейки позволяет реализовать матрицу запоминающего устройства с побитным доступом и с последовательным выводом данных через общую шину. Для этого в топологии, показанной на рис.10в, объединены ячейки в кроссбар по внешним линиям комплементарной пары мемристоров. В новой топологии изменено расположение одной из верхних линий, подключенной к мемристорам, с вертикального прохождения на горизонтальное. При этом полученное пересечение проводников разделено в пространстве слоем диэлектрика. Для соединения кросспроводника с нижними мемристорами сформированы проводящие переходные колодцы. Линии электрической связи диодных катодов в ячейках соединили в один проводящий слой, являющейся общей шиной последовательного вывода данных. Роль катодного слоя может играть подложка легированного полупроводника р-типа.

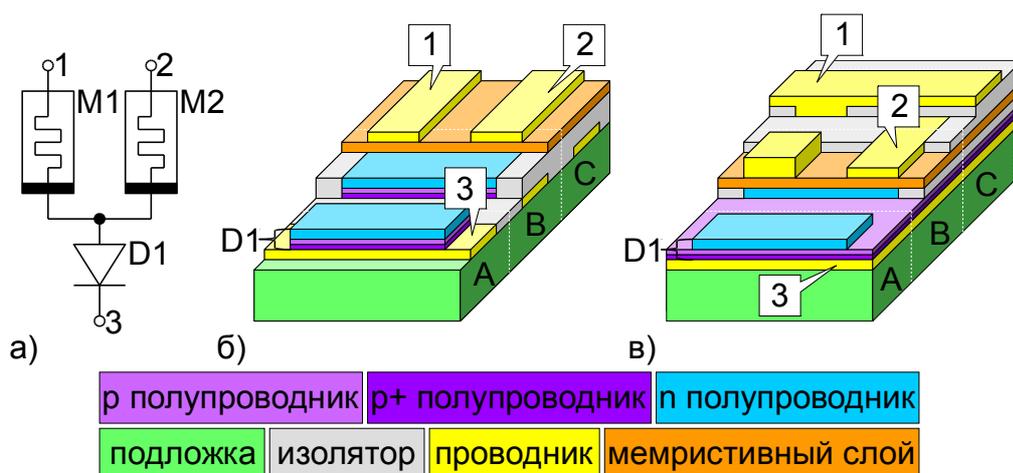


Рис.10. Схема комплементарной мемристивно-диодной ячейки а); топология ячейки для матрицы: с параллельным выводом информации б) и с последовательным выводом информации в).

Следует отметить, что ячейка для матрицы с последовательным выводом информации технологически более сложная. Она требует нанесения верхнего слоя диэлектрика и имеет дополнительную технологическую трудоемкость при создании переходных проводящих колодцев. Однако в этой ячейке не требуется литография для создания проводников катодов, и соединение катодов осуществляется с помощью одного проводящего слоя легированного акцепторной примесью полупроводника. Преимуществом более сложной ячейки является значительное уменьшение межшинной емкости, что увеличивает энергоэффективность при работе матрицы на высоких скоростях записи.

2.5. Топология 3D матрицы на основе интеграции двухслойного мемристорно-диодного кроссбара.

Высокая степень интеграции запоминающей матрицы достигается за счёт трехмерной компоновки элементов, образующих 3D структуру [76]. Эта структура состоит из горизонтально расположенных и зеркально ориентированных по отношению друг к другу одинаковых комбинированных кроссбаров, включающих ячейки 1D2M. Соединение между кроссбарами осуществляется по общим шинам строк и столбцов. Два соседних кроссбара с общими шинами, соединёнными с катодами диодов Зенера, образует отдельный функциональный пласт. Все функциональные пласти идентичны, накладываются друг на друга и объединяются общими шинами в электрическую цепь. Принцип объединения комбинированных кроссбаров в электрическую цепь показана на рис.11 [77]. Электроды ячеек Y1 и Y2 соединены с катодами диода Зенера. Анодом диода Зенера является металлический проводник непосредственно соприкасающийся с вышележащим сплошным мемристорным слоем. С другой стороны мемристорного слоя расположены верхние электроды элементарной ячейки X1-X4. Два соседних перекрестья металлических проводников образуют комплементарные мемристоры.

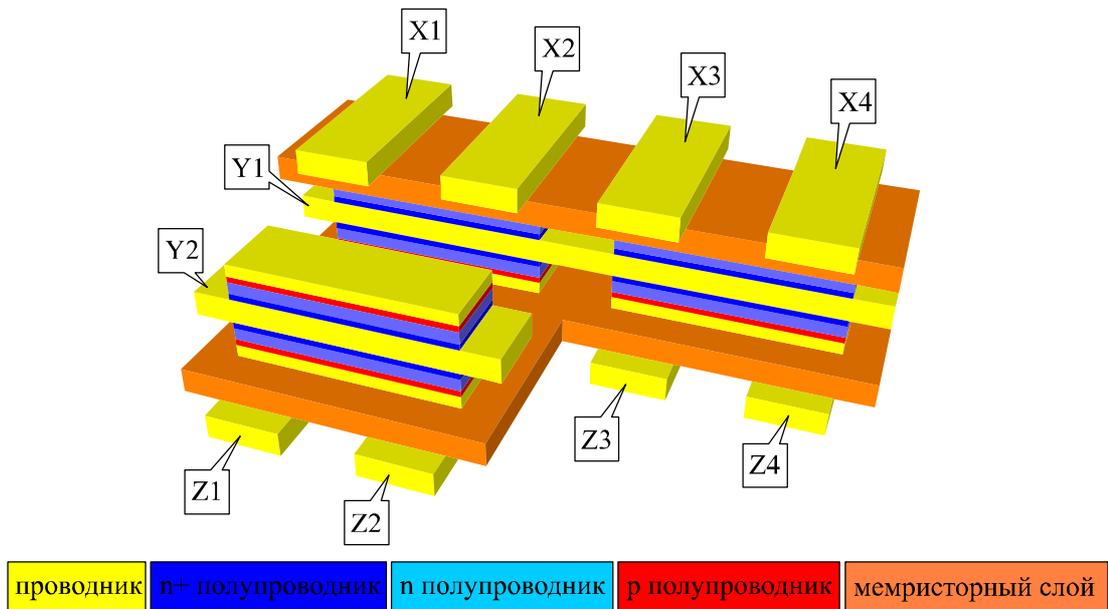


Рис.11. Принцип объединения комбинированных кроссбаров в 3D запоминающую матрицы с высокой интеграцией элементов.

На рис. 12. представлена топология фрагмента запоминающей матрицы, состоящего из трех горизонтально расположенных и зеркально ориентированных по отношению друг к другу кроссбаров, в каждом из которых находится 18 ячеек.

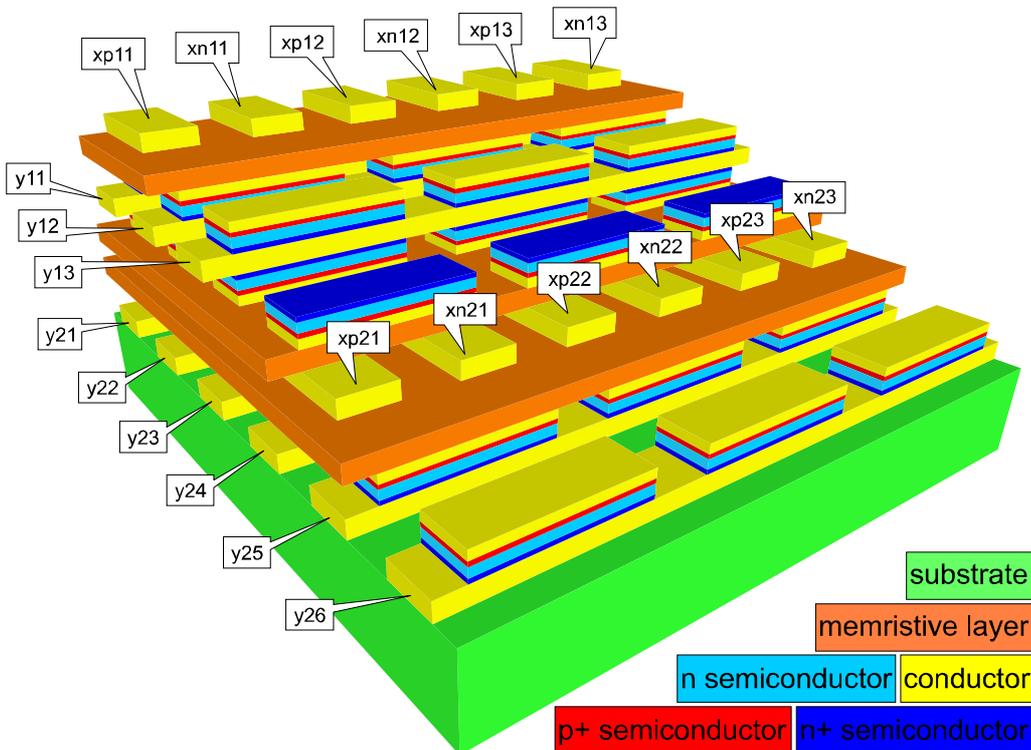


Рис.12. Топология фрагмента 3D запоминающей матрицы.

Топология всей трехмерной матрицы выстраивается путем последовательного наращивания следующего функционального пласта с инвертированным порядком изготовления слоев, при этом проводники строк или столбцов у соседних кроссбаров будут общими.

2.6. Выводы к главе 2.

Разработана электрическая схема и топология ячейки запоминающей матрицы, состоящей из комплементарной пары мемристоров и селективного диода Зенера.

Комплементарное включение двух мемристоров позволяет постоянно поддерживать высокое входное сопротивление ячейки в рабочем режиме, что обеспечивает ее низкое энергопотребление. Поскольку комплементарные мемристоры пространственно находятся в одном активном слое рядом друг с другом, то влияние неоднородности их характеристик на выходное напряжение ячейки будет минимальным.

Использование диода Зенера, имеющего низкое напряжение обратимого пробоя, устраняет проблему возникновения паразитных токов при работе запоминающей матрицы в импульсном режиме и увеличивает ее размер, а также позволяет использовать биполярные мемристоры.

На основе двух топологий комплементарной мемристорно-диодной ячейки разработаны два варианта запоминающей матрицы – с параллельным и последовательным выводом данных. Предложенные топологии мемристорно-диодных ячеек дают возможность добиться высокой степени интеграции при объединении их в сверхбольшую матрицу, в которой крупные КМОП-транзисторы являются общими для больших строк ячеек.

Высокая степень интеграции запоминающей матрицы достигается за счёт трехмерной компоновки элементов, образующих 3D структуру. Эта структура состоит из горизонтально расположенных и зеркально ориентированных по отношению друг к другу одинаковых комбинированных кроссбаров, включающих ячейки 1D2M.

ГЛАВА 3. SPICE - моделирование работы сверхбольшой запоминающей матрицы.

Название компьютерной программы SPICE это акроним для Simulation Program with Integrated Circuit Emphasis – программа моделирования с акцентом на интегральные схемы. Она была разработана в середине 70х годов в университете Беркли в Калифорнии, для моделирования интегральных электрических схем перед производством. До ее появления тестирование схем производилось на дорогостоящих опытных образцах. В наше время SPICE совместимые программы являются стандартом для моделирования на этапе проектирования как аналоговых, так и цифровых устройств.

Электрическая схема в SPICE задается как набор элементов и связей между ними. Такой список можно составить вручную или с помощью графического интерфейса. Базовый набор элементов включает резисторы, конденсаторы, индуктивные катушки, диоды, биполярный и полевые транзисторы, длинные линии, источники напряжения и тока различной формы, а также элементы цифровой логики.

Упрощенная блок схема алгоритма работы программы SPICE приведена на рисунке 9. Алгоритм состоит из трех основных действий:

1. Основная часть – анализ узлов (блоки 3 и 4) формирует узловую матрицу на основе законов Кирхгофа и решает уравнения для напряжений в цепях.

2. Внутренний цикл 2-6 находит решение для нелинейных цепей методом итераций. Нелинейные элементы при этом заменяются эквивалентными линейными моделями.

3. Внешний цикл 7-9 вместе с внутренним выполняет анализ переходных процессов, заменяя запасующие энергию элементы (емкости и индуктивности) эквивалентными линейными моделями, а также подбирает временные интервалы.

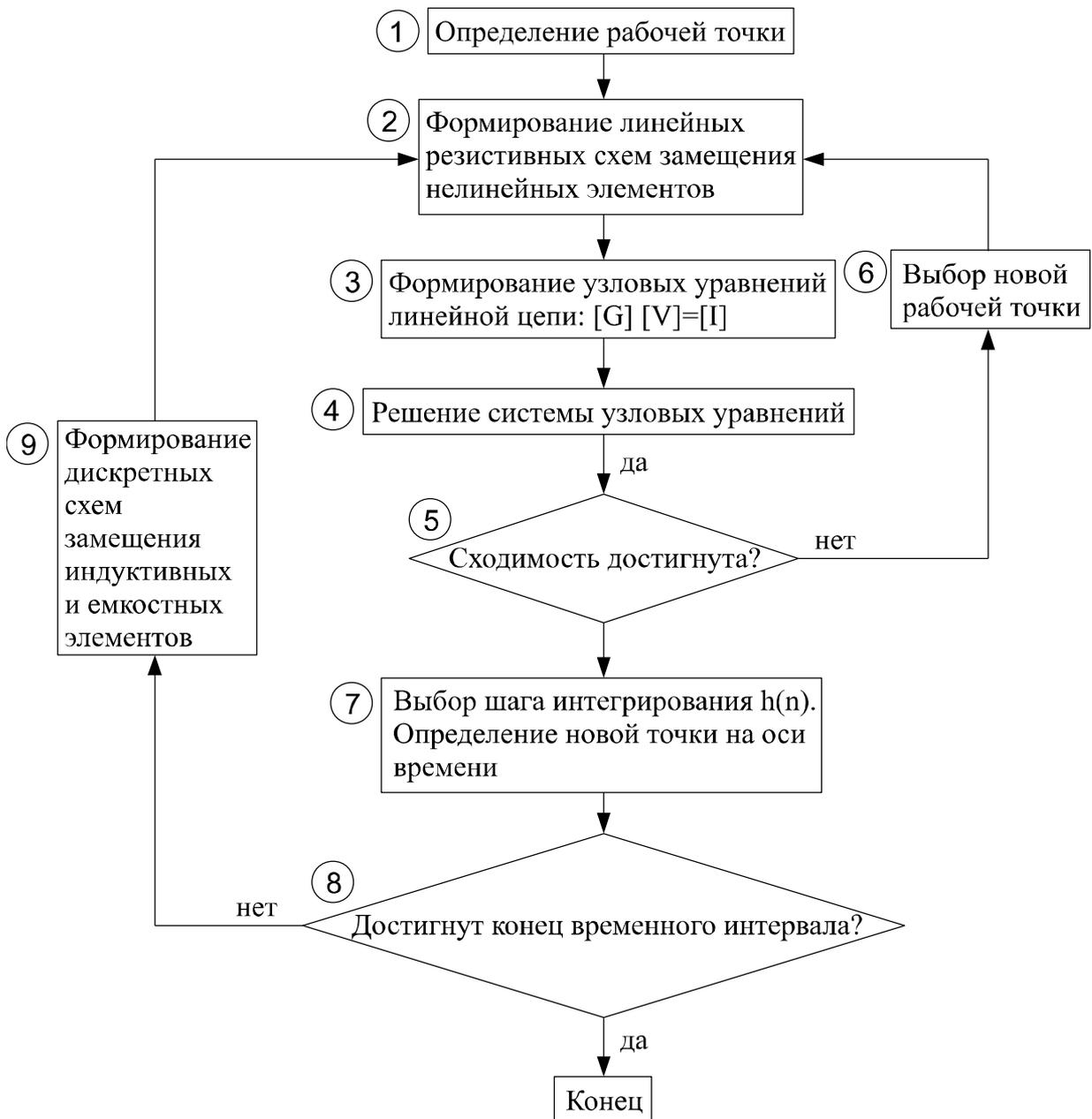


Рис.13. Блок схема расчета программы SPICE.

Цифрами обозначены операции.

Таким образом SPICE может выполнять три основных вида анализа электрических цепей:

1. Анализ нелинейных резистивных цепей постоянного тока (DC analysis);
2. Анализ частотного отклика линейных цепей (AC analysis);
3. Анализ переходных процессов нелинейных динамических цепей при воздействии сигналов произвольной формы (transient analysis).

Анализ цепей постоянного тока применяется как непосредственно для вычисления реакции при подаче постоянного напряжения, так и при определении рабочей точки нелинейной цепи. Рабочая точка является начальным решением для переходных процессов и точкой смещения при частотном отклике. В этом режиме конденсаторы считаются разрывом цепи, выводы индуктивности шунтируются.

Простейший случай в данном режиме – анализ линейных резистивных цепей. При этом используются операции 3 и 4 из алгоритма (см. рис. 13), выполняющие, соответственно, формирование матрицы алгебраических уравнений и ее решение методом Гаусса. Для анализа линейной резистивной цепи используется модифицированный метод узловых напряжений.

При анализе нелинейных резистивных цепей задействуется внутренний цикл 1-6, реализующий итеративные вычисления по методу Ньютона-Рафсона. На каждой итерации формируются узловые уравнения эквивалентной линейной резистивной цепи. Результатом работы итеративной процедуры является более близкое к истинному значение рабочей точки.

Анализ частотного отклика состоит из двух частей. Первая часть выполняет расчет рабочей точки для нелинейных цепей (операции 1-6), для постоянного тока. Во второй части нелинейные компоненты заменяются линеаризованными моделями с параметрами, соответствующими рабочей точке. После этого выполняется решение узловой матрицы в комплексной форме для заданных частот. Если в цепи действуют несколько синусоидальных источников, их частоты полагаются одинаковыми. При расчете частотных характеристик анализируется линеаризованная схема замещения, искажения сигнала, обусловленные нелинейностью ВАХ элементов, отсутствуют. Поэтому токи и напряжения могут достигать очень больших величин, не соответствующих действительности.

Для расчета переходных процессов используются неявные методы численного интегрирования – методы трапеций, Гира второго порядка или неявный метод Эйлера. По умолчанию используется метод трапеций.

Максимальный шаг интегрирования выбирается пользователем или устанавливается автоматически. SPICE автоматически выбирает шаг интегрирования, обеспечивающий оптимальное соотношение между точностью и временем расчета.

Процедура расчета переходного процесса является многошаговой. На каждом шаге интегрирования автоматически определяется рабочая точка – токи и напряжения нелинейных компонентов. При определении рабочей точки нелинейной цепи напряжения и токи источников сигнала полагаются равными нулю, индуктивные элементы заменяются коротким замыканием, а емкостные – разрывом. Расчет рабочей точки ведется итеративным методом Ньютона-Рафсона. На каждой итерации нелинейные компоненты заменяются линеаризованными схемами замещения, соответствующими режиму этого компонента.

Анализ нелинейных динамических цепей является наиболее сложным видом анализа. В этом режиме используются все модули алгоритма.

3.1. Физико-математическая модель работы ячейки матрицы.

С целью создания сверхбольшой запоминающей матрицы создана физико-математическая модель работы ячейки на основе упрощенных моделей мемристора и диода Зенера. В модели мемристора изменение параметра состояния было жестко ограничено по сравнению с [79], поскольку неабсолютная точность рациональных чисел в компьютерной системе приводит к выходу параметра состояния за границы допустимого интервала. Для диода Зенера построена идеализированная модель, нелинейная вольт-амперная характеристика которой представляет собой кусочную функцию из трех прямых линий [69], в которой сопротивление остается высоким в диапазоне от напряжения обратимого пробоя до напряжения открытия p-n перехода.

Выбор в пользу этой модели диода сделан на основании того, что детальная макроскопическая модель [58] при совпадении вольтамперных

характеристик (рис.14а) обладает существенным недостатком, связанным с неправильным моделированием сопротивления утечки (рис.14б).

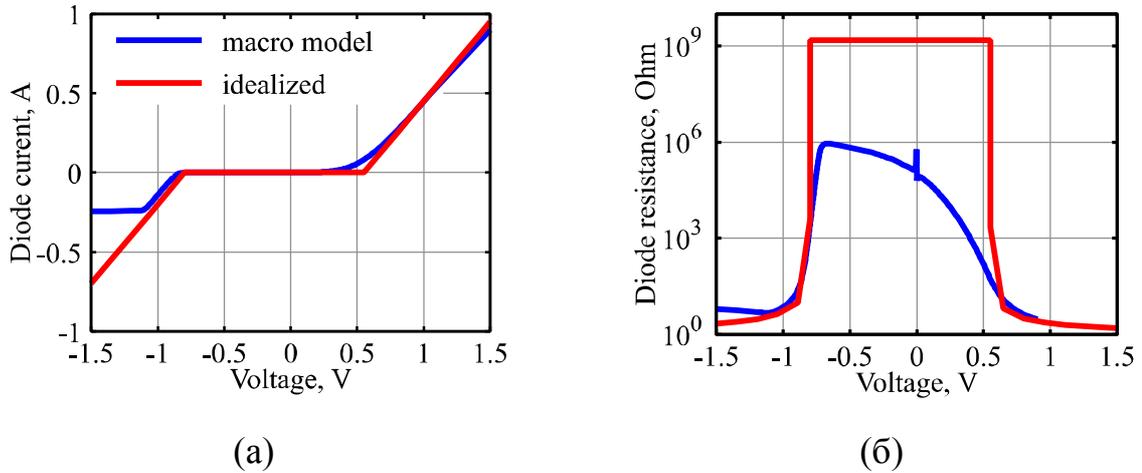


Рис.14. Сравнение идеализированной (красная кривая) и детальной (синяя кривая) моделей диода Зенера: (а) вольтамперные характеристики, (б) сопротивление диода Зенера от напряжения на его контактах.

Модель биполярного мемристора учитывает нелинейный ионный дрейф и пороговый характер переключения [79]. В этой модели мемристивность (сопротивление мемристора) R выступает в качестве параметра состояния x ($x \equiv R$). Состояние системы описывается следующими уравнениями.

$$I = \frac{V}{x}, \quad \frac{dx}{dt} = f(V)W(x, V), \quad (1)$$

где f – кусочная функция, определяющая пороговое переключение, а W – оконная функция, описывающая характер и пределы изменения сопротивления, V – напряжение на контактах мемристора.

$$f(V) = \beta(V - 0,5(|V + V_t| - |V - V_t|)), \quad (2)$$

$$W(x, V) = H(V)H(R_{off} - x) + H(-V)H(x - R_{on}).$$

Здесь H – функция Хевисайда, либо сглаженная ступенька, β – константа, определяющая скорость изменения параметра состояния при $|V| > V_t$.

Согласно определению мемристивной системы изменение параметра x состояния во времени определяется функцией f . Соответственно, для вычисления самого параметра состояния необходимо произвести интегрирование этой функции. Для выполнения этой операции в программе SPICE создана электрическая схема, выполняющая интегрирование. Обычно для этих целей используется конденсатор с регулируемым источником тока Gx . Мемристор в этом случае представлен как источник тока Gm , величина которого зависит от напряжения на контактах устройства и внутреннего параметра состояния (рис. 15).

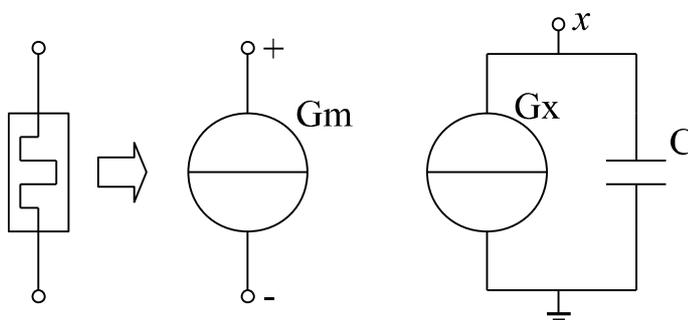


Рис.15. Общая эквивалентная электрическая подсхема мемристора.

Величина тока Gx определяется скоростью изменения параметра состояния x . Конденсатор C за шаг моделирования Δt заряжается источником тока до напряжения:

$$V_c = \frac{1}{C} \int_t^{t+\Delta t} I(Gx) dt + V_c(t). \quad (3)$$

Если задать емкость конденсатора равной 1 Ф, то напряжение на выводе x на рис.15 будет равно интегралу от тока источника Gx .

Ток источника Gm вычисляется непосредственно как функция $g(V_m, x)$ напряжения на контактах мемристора V_m и вычисленного ранее параметра состояния x , представленного напряжением на конденсаторе.

При необходимости хранения каких-либо переменных значений в подсхему можно ввести дополнительные источники напряжения, у которых можно непосредственно в SPICE коде менять потенциал на выходе.

Функция Хевисайда $H(x)$ имеет разрыв в точке $x = 0$, что может привести к ситуации когда численное решение не будет сходиться. Чтобы этого избежать в работе [79] предложена сглаженная функция Хевисайда.

$$\theta(x) = \frac{1}{1 + \exp\left(-\frac{x}{b}\right)}. \quad (4)$$

Константа b определяет скорость изменения функции и $\theta(x) \rightarrow H(x)$ при $b \rightarrow \infty$. Первая производная модуля также претерпевает разрыв в точке $x = 0$. Непрерывную функцию модуля можно определить на основе функции $\theta(x)$:

$$\text{abs}(x) = x[\theta(x) - \theta(-x)]. \quad (5)$$

SPICE код модели:

```
*Мемристор с пороговым переключением
.subckt memR_TH plus minus PARAMS: Ron=10K Roff=100K
+Rinit=99K beta=1E13 Vt=4.6

*Внешний интерфейс мемристора
Gpm plus minus value={V(plus,minus)/V(x)}
* Интегратор
Evar var 0 value={fs(V(plus,minus),b1)*ws(v(x),V(plus,minus),b1,b2)*1p}
Gx 0 x value={IF(V(x)>Ron,
+      IF(V(x)<Roff,
+      V(var),
+      IF(V(plus, minus)<0,
+      V(var),
+      0)),
+      IF(V(plus, minus)<0,0,V(var))
+      )
+      }
Cx x 0 1p IC={Rinit}
*Сглаженные функции (без разрывов)
.param b1=10u b2=10u
.func stps(x,b)={1/(1+exp(-x/b))}
.func abss(x,b)={x*(stps(x,b)-stps(-x,b))}
.func fs(v,b)={beta*(v-0.5*(abss(v+Vt,b)-abss(v-Vt,b)))}
.func ws(x,v,b1,b2)={stps(v,b1)* stps(1-x/Roff,b2)+stps(-v,b1)*
+stps(x/Ron-1,b2)}

.ends memR_TH
```

В качестве примера на рис.16 приведены вольтамперные характеристики модели мемристора, полученные при подаче на него синусоидального 5 В напряжения с разной частотой (порог переключения в модели $V_t = 4,6$ В).

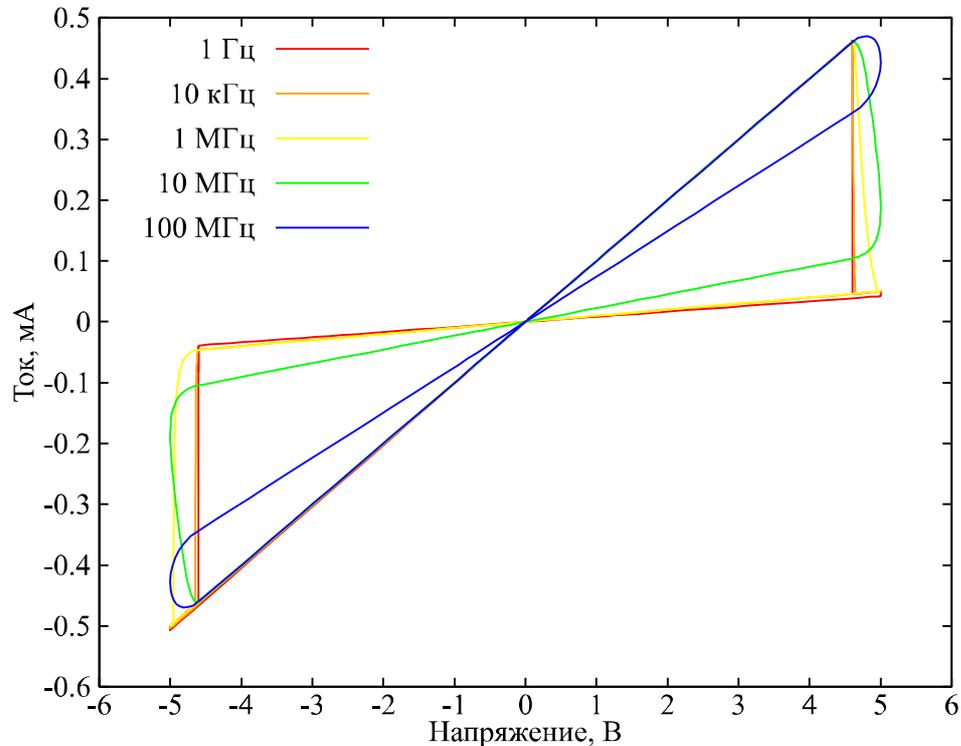


Рис.16. Семейство вольтамперных характеристик мемристора с пороговым переключением.

Семейство вольтамперных характеристик модели с пороговым переключением удовлетворяет трем признакам мемристора: петля гистерезиса имеет двойную точку в начале координат, площадь петли гистерезиса монотонно уменьшается с ростом частоты входного сигнала, петля гистерезиса вырождается в линию при стремлении частоты входного сигнала к бесконечности. Модель корректно работает в большом диапазоне частот.

Результат записи в первую ячейку, полученный в ходе SPICE моделирования работы матрицы из 2 x 2 ячеек, представлен на рис.17 [68], из которого видно, что комплементарные мемристоры при считывании всегда

находятся в противоположных состояниях. Изначально все мемристоры находились в промежуточном состоянии 500 кОм. При записи в первую ячейку сопротивление одного мемристора M1 (синяя кривая) увеличивается до предела (1МОм), а второго M2 (оранжевая кривая) уменьшается до минимального (10 кОм). Таким образом, общее сопротивление пары остается высоким.

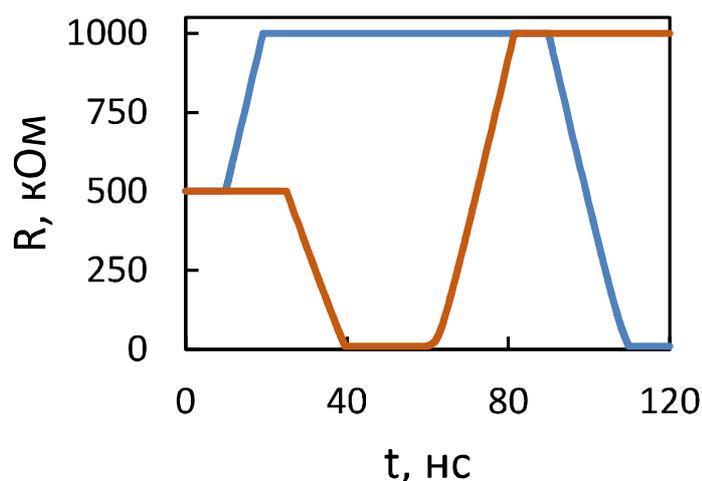


Рис.17. Кривая изменения сопротивлений M1 (синий) и M2 (оранжевый) во времени.

Процесс изменения состояния ячейки производился путем последовательного изменения сопротивления двух мемристоров. Численный расчет с использованием разработанной модели ячейки запоминающей матрицы показывает (рис.18), что запись одной ячейки из комплементарных мемристоров в матрице размером 100x100 требует в 8 раз меньше энергии при добавлении в каждую ячейку диода Зенера, значительно уменьшающего паразитные токи через соседние ячейки. Для моделирования взят худший возможный при записи случай, когда все ячейки матрицы изначально находились в одинаковом состоянии. Отношение сопротивлений мемристора в низкопроводящем и высокопроводящем состояниях $R = (R_{off} - R_{on})/R_{on} = 10$.

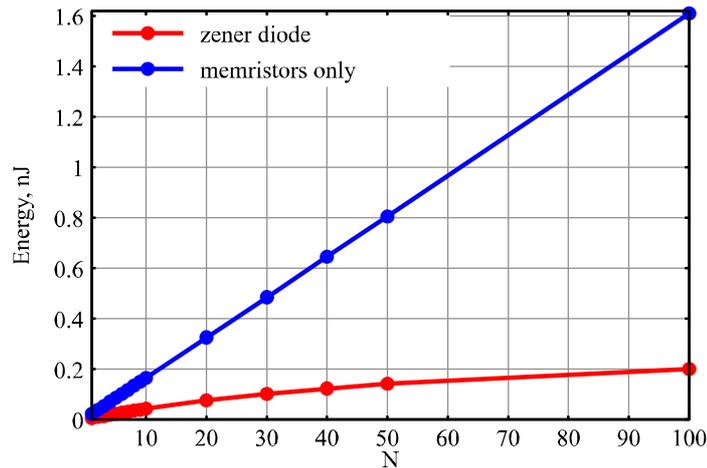


Рис.18. Затраты энергии на запись одной ячейки комплементарных мемристоров без диода (синий цвет) и с диодом Зенера (красный цвет) в зависимости от числа ячеек в квадратной матрице $N \times N$.

В сверхбольшой матрице, предложенной в настоящей работе, производится взвешивание и суммирование отдельных импульсов. Работу с отдельными входными импульсами можно рассматривать как последовательное считывание. Схема подачи входных импульсов описана в параграфе 3. Рис. 19 показывает деградацию выходного напряжения от размера квадратной матрицы $N \times N$ при подаче одного импульса амплитудой 1 В для R равным 10, 100 и 1000.

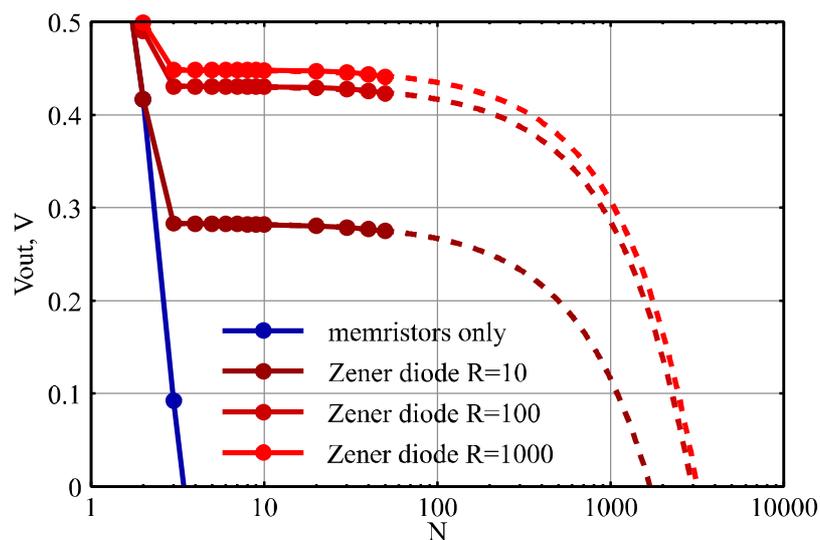


Рис.19. Зависимость выходного напряжения от размера матрицы при последовательном чтении при разных R .

Из рис.19 видно, что при отсутствии в ячейках диода Зенера выходное напряжение снижается практически до нуля уже в 3x3 матрице. При добавлении диода Зенера происходит снижение выходного напряжения в пределах от 50% до 70% в матрице того же размера, а дальнейшее увеличение размера матрицы слабо влияет на величину выходного сигнала. Медленно изменяющийся уровень выходного напряжения $\sim 0,3$ В достаточен для выполнения дальнейшей процедуры суммирования.

Следовательно, матрица с комплементарными мемристорами и без нелинейного селективного элемента [39] не может быть использована в качестве сверхбольшой матрицы нейропроцессора.

3.2. Физико-математическая модель работы запоминающей матрицы.

В запоминающей матрице умножение входного сигнала на вес связи выполняется с использованием закона Ома. Напряжение на входе сумматора при приложении к контактам мемристоров ячейки напряжения разной полярности амплитудой U_{input} имеет вид [69]:

$$U_i = \left(\frac{2U_{input}}{R_{1i} + R_{2i}} R_{2i} - U_{input} \right) \frac{1}{R_D + R_P} R_P, \quad (6)$$

где R_{1i} и R_{2i} – сопротивления мемристоров комплементарной пары, R_D – сопротивление прямосмещенного диода Зенера, R_P – входное сопротивление сумматора. Сопротивление мемристоров представляет собой континуум значений в пределах от минимального R_{on} до максимального R_{off} .

Так как мемристоры образуют комплементарную пару, сумма их сопротивление остается постоянной: $R_{1i} + R_{2i} = K = const$. Учитывая, что сопротивление открытого диода мало, $R_D \ll R_P$, то справедливо следующие выражения для напряжений на входе и выходе сумматора с высоким входным сопротивлением соответственно:

$$U_i = \frac{U_{input}}{K} (2R_{1i} - K), \quad (7)$$

$$U_s = \sum_i U_i = \frac{U_{input}}{K} \sum_i (2R_{1i} - K). \quad (8)$$

Суммирование напряжений выполняется последовательной подачей напряжения на выбранные ячейки.

3.3. Численное моделирование процессов обработки сигналов при взвешивании импульсов напряжения и сложении токов ячеек в запоминающей матрице и сравнение с экспериментальными данными.

В качестве сумматора взята простая схема из КМОП инвертора и конденсатора (рис.20) [69]. Сильная нелинейность передаточной характеристики инвертора, образованного транзисторами T1 и T2, не позволяет задействовать промежуточные состояния мемристоров. При разработке нейропроцессора включение промежуточных состояний мемристоров можно обеспечить модернизацией схемы сумматора, в котором ток заряда конденсатора будет пропорционален сопротивлению мемристора.

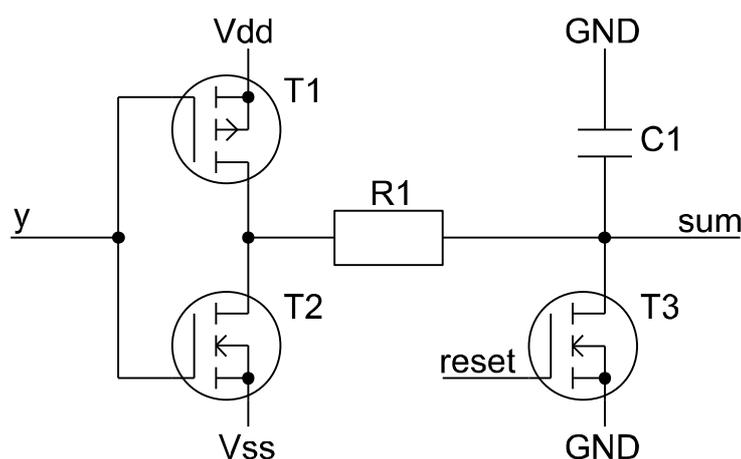


Рис. 20. Схема сумматора. Vdd и Vss – полюса питания.

GND – шина с нулевым потенциалом.

Проверка работоспособности фрагмента матрицы заключалась в организации трех состояний ячеек: оба мемристора закрыты ($R_1 = R_{off}, R_2 =$

R_{off}), один из мемристоров пары открыт ($R_1 = R_{on}, R_2 = R_{off}$ и $R_1 = R_{off}, R_2 = R_{on}$), а также взвешенного суммирования входных сигналов, включающего сложение и вычитание. На рис.21 показано изменение во времени выходного напряжения сумматора V_{out} .

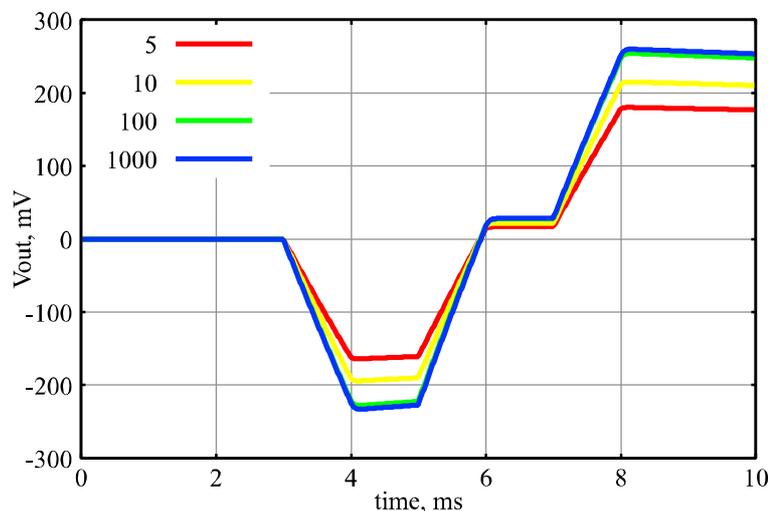


Рис.21. Выходное напряжение сумматора.

Цвет соответствует различным R .

SPICE-модель позволяет оптимизировать параметры элементов ячейки и матрицы в целом. Из рисунка видно, что чем выше R , тем сильнее состояние ячейки влияет на выходное напряжение сумматора.

На быстродействие матрицы будут влиять паразитные емкости шин, мемристоров, диодов и в первую очередь входная емкость усилителя. Время установления напряжения на выходной шине в матрицах размером от 1x1 до 50x50 без учета емкости усилителя практически не меняется и равно 41 пс. При моделировании учитывалась емкость проводников (толщиной 30 нм и с межшинным расстоянием 200нм), диодов (площадь 600x200 нм²) и мемристоров (площадь 200x200 нм², толщина 30 нм). Моделирование матрицы с моделью инвертора РТМ 90nm [80] дает задержку 5,62 нс.

Сложение токов, протекающих через закрытые диоды Зенера, осуществляется при низком входном сопротивлении усилителя. Сумма выходных токов ячеек I_s :

$$I_s = \sum_i \left(\frac{2}{K} R_{2i} - 1 \right) \frac{U_{input}}{R_D} \quad (9)$$

Взвешивание напряжений входных импульсов с последующим суммированием токов позволяет уменьшить паразитные токи между ячейками и произвести корректное сложение одновременно пришедших импульсов. На рис.22. показан выходной ток одной шины матрицы как результат сложения токов из двух ячеек, находящихся в разных синаптических состояниях [1].

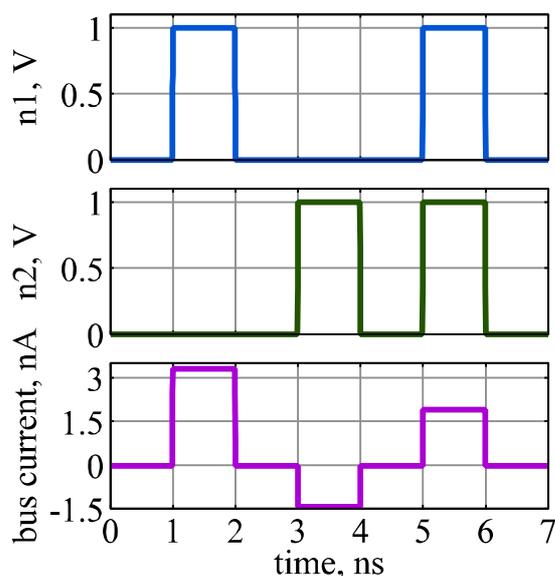


Рис.22. Взвешенное суммирование входных импульсов напряжения: при низком входном сопротивлении сумматора.

SPICE-модели позволяют оптимизировать параметры элементов ячейки и матрицы в целом. Из рисунка видно, что чем выше отношение R сопротивлений мемристоров в низкопроводящем и высокопроводящем состояниях, тем сильнее состояние ячейки влияет на выходное напряжение сумматора.

При моделировании матричного умножения входного вектора на матрицу чисел [81, 82] на входные проводники кроссбара запоминающей матрицы поступают импульсы напряжения. Выходные токи кроссбара являются суммой токов от отдельных ячеек, которые в свою очередь однозначно определяются значениями сопротивления мемристоров. На

рис.23 показаны выходные токи кроссбара размером 2x2 при трех разных состояниях весовых коэффициентов. Выходной ток с первого выходного проводника обозначен как Out[0], а второй – Out[1]. Для сравнения приводятся экспериментальные данные по матричному умножению из [83, 84].

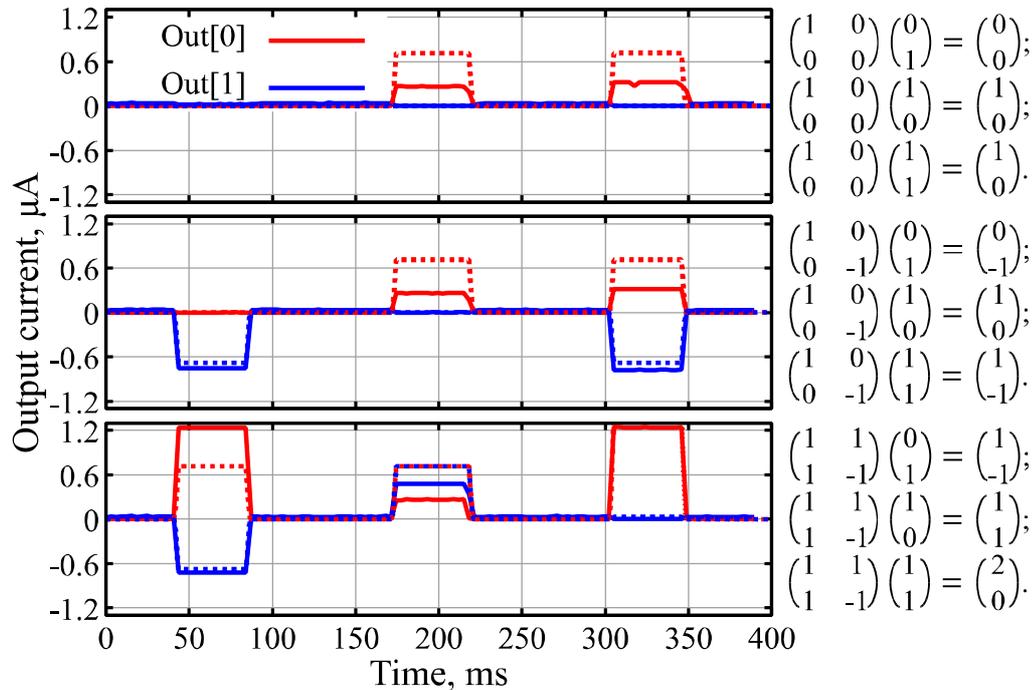


Рис.23. Результат матричного умножения матрицы чисел размером 2x2 на двухкомпонентный вектор: сплошные кривые – экспериментальные, пунктирные кривые - SPICE-моделирование.

Результаты моделирования показывают, что полученный выходной вектор токов соответствует результату матричного умножения матрицы на вектор, полученному по математическим правилам.

Из-за высокого коэффициента усиления в преобразователе ток-напряжение и превышения максимального входного напряжения АЦП в микроконтроллере ожидаемый высокий ток на Out[0] в последнем случае не наблюдается. Среднеквадратичное отклонение выходных токов за время действия входных импульсов напряжения, вносимая неодинаковостью мемристоров, оценена при сравнении экспериментальных данных и SPICE-моделирования и составляет 276 нА.

3.4. Сравнение результатов численного моделирования работы запоминающей матрицы с известными аналогами

Включение в комплементарную ячейку диода Зенера может уменьшить деградацию выходного сигнала при параллельном считывании в запоминающей матрице [32] с большим количеством ячеек, предназначенной только для хранения двоичной информации. Рис.24 иллюстрирует зависимость величины разности токов считывания в двух плечах Pre-Charged Sense Amplifier (PCSA) усилителя, используемого в [32], от размера кроссбара при $R = 100$.

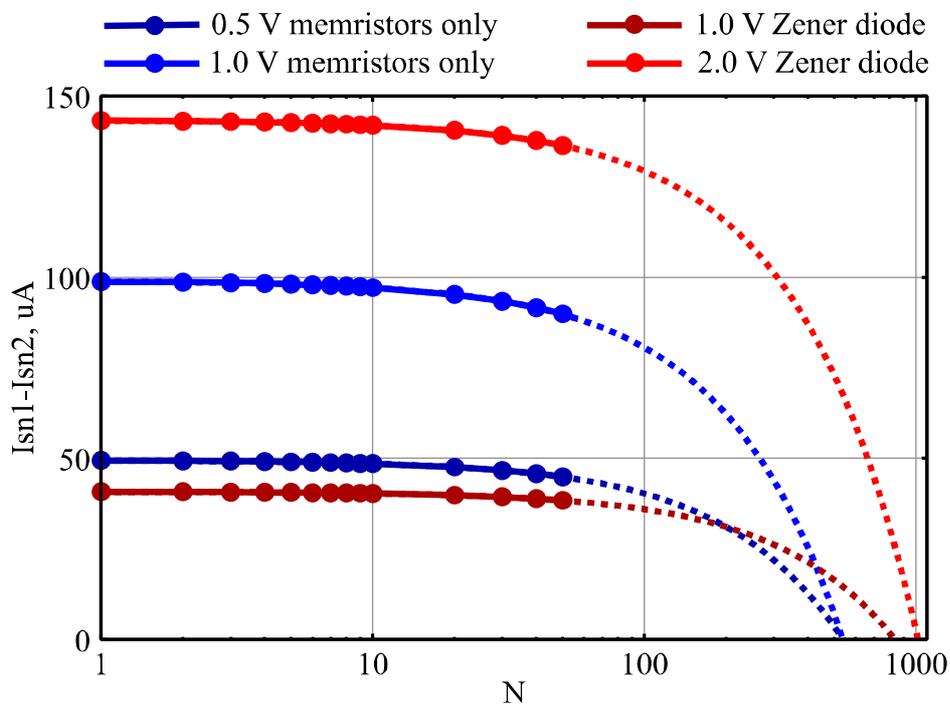


Рис.24. Разница токов считывания в двух плечах (I_{sn1} и I_{sn2}) PCSA усилителя в зависимости от размера квадратной матрицы $N \times N$ ячеек.

Как видно из рисунка разность токов стремится к нулю с увеличением числа ячеек любом напряжении питания, что приводит неработоспособности усилителя.

Схема ввода входных импульсов в матрицу комплементарных мемристоров без диодов, предложенная в [39], предполагает подачу входного импульса на общий контакт мемристоров. Выходной сигнал,

представляющий собой разницу напряжений на остальных двух контактах ячейки, передается на сумматор по шинам, объединяющим ячейки в строку. На рис. 25 представлены результаты SPICE моделирования процесса прохождения одного импульса с амплитудой 1 В (считывания состояния одной ячейки) по этой схеме.

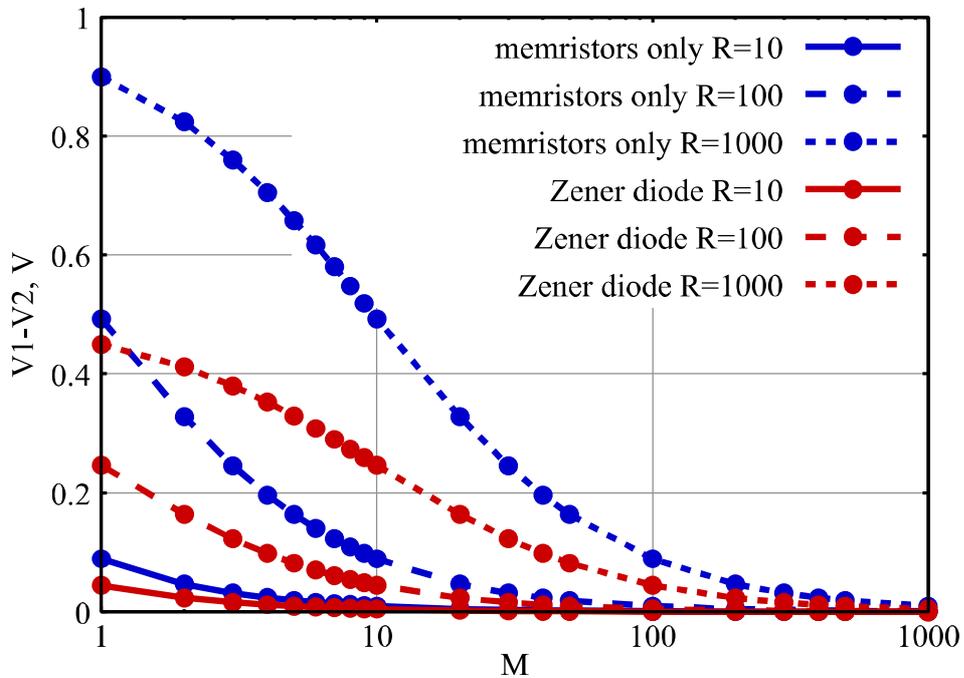


Рис.25. Зависимость выходного дифференциального напряжения в зависимости от размера матрицы при последовательном чтении при разных R .

Результат моделирования показывает, что при такой схеме подачи входных импульсов происходит практически полная деградация выходного сигнала при малых размерах матрицы как с диодом Зенера, так и без него.

Следовательно, матрица с комплементарными мемристорами и без нелинейного селективного элемента [39] не может быть использована в качестве сверхбольшой матрицы нейропроцессора.

Высокая степень интеграции подтверждается сравнением в таблице 1, в которой даны площади одной ячейки на основе мемристоров в представленной и известных запоминающих матрицах.

Таблица 1. Площадь ячейки при различном включении мемристоров.

| Структура синапса | Площадь, F ² | Ссылка |
|-------------------|-------------------------|----------------------|
| 2M | 8 | [39] |
| 1D2M | 8 | Разработанная ячейка |
| 2×1T1M | 24.8 | [40] |
| 2T1M | 40.3 | [4] |

Площадь ячейки матриц [4] и [40] определяется суммарной площадью транзисторов, поскольку предполагается, что мемристор расположен над ними. Для оценки был использован размер планарных КМОП транзисторов из работы [46].

3.5. Выводы к главе 3

Создана физико-математическая модель работы ячейки и соответствующая SPICE-модель на основе упрощенных моделей мемристора и диода Зенера с целью расчета кроссбаров большого размера. Для диода Зенера построена идеализированная модель, нелинейная вольт-амперная характеристика которой представляет собой кусочную функцию из трех прямых линий. Модель биполярного мемристора учитывает нелинейный ионный дрейф и пороговый характер переключения.

Численный расчет с использованием разработанной модели ячейки запоминающей матрицы показывает, что запись одной ячейки из комплементарных мемристоров с $(R_{off} - R_{on})/R_{on} = 10$ в матрице размером 100x100 требует в 8 раз меньше энергии при добавлении в каждую ячейку диода Зенера. Деграция входного напряжения ячейки в запоминающей матрице достигает 30% от первоначального (1 В) при размере матрицы равном 1000x1000 ячеек.

Описаны и численно промоделированы процессы умножения входного напряжения на весовой коэффициент синаптической связи в ячейке, сложение получившихся напряжений во времени при использовании сумматора с высоким входным сопротивлением и сложение получившихся токов при использовании сумматора с низким входным сопротивлением. С помощью численного моделирования проверена работа запоминающей матрицы размера 2×2 при взвешенном суммировании импульсов напряжения. Среднеквадратичное отклонение выходных токов за время действия входных импульсов напряжения, вносимая неодинаковостью мемристоров, оценена при сравнении экспериментальных данных и SPICE-моделирования и составляет 276 нА.

При сравнении разработанной электрической схемы запоминающей матрицы с известными аналогами показано, что она обладает наименьшей площадью ячейки и может обладать большим размером из-за меньшего влияния паразитных токов.

ГЛАВА 4. Исследование работоспособности запоминающей матрицы в составе аппаратной импульсной нейросети

4.1. Использование запоминающей матрицы в качестве матрицы синапсов аппаратной импульсной нейросети.

Большая часть расчета нейросети по взвешиванию и суммированию реализуется в аппаратном виде следующим образом. Входной сигнал представлен в виде двух напряжений противоположной полярности и подается на контакты резистивного делителя напряжения, при этом напряжение в средней точке будет определяться соотношением сопротивлений резисторов. В такой схеме реализуется умножение амплитуды входного напряжения на величину, однозначно определяемую сопротивлениями. Применение мемристоров позволяет устанавливать эту величину произвольно.

Математическая функция, выполняемая формальным нейроном, имеет вид [69]:

$$y_j = f_{act} \left(\sum_i w_{ij} x_i \right), \quad (10)$$

где f_{act} – функция активации, а в ее аргументе x_i – сигналы от предыдущих нейронов, w_i – вес связи данного нейрона с i -тым нейроном.

Таким образом, соотношение электрических величин в аппаратной части, описываемое выражением (4), соответствует аргументу активационной функции в формуле (1) следующим образом:

$$x_i \rightarrow U_{input}, \quad w_i \rightarrow \frac{2R_{1i} - K}{K}.$$

Сигнал U_{input} на входные шины подается импульсами, причем на один вход ячейки импульс приходит неизменным, а на другой – в инвертированном виде. Импульсы на общей шине создают колебания потенциала U_i , которые поступают на вход сумматора, например, построенного по схеме классического интегратора, на выходе которого уровень напряжения U_S определяется частотой поступающих импульсов на вход матрицы. Таким

образом, импульсы напряжения в комплементарной паре мемристоров, работающих как делитель напряжения, взвешиваются аналогично синаптическому взвешиванию в нейронной сети, и в зависимости от записанных в мемристорах проводимостях поступают в сумматор.

В качестве проверки возможности использования разработанной запоминающей матрицы в качестве массива синапсов аппаратной импульсной нейросети построена сеть прямого распространения информации – однослойный аппаратный персептрон [85, 86]. Сеть выполняет распознавание цифр из датасета MNIST, содержащего набор рукописных цифр от 0 до 9. Исходное разрешение входных изображений было понижено с 28×28 до 7×7 с целью уменьшения сложности расчетов и времени моделирования. Соответственно сеть имеет 49 входных нейронов, 10 выходных нейронов и 490 синапсов. При расчете одного прохода каждый входной нейрон выдает последовательность импульсов со средней частотой, пропорциональной яркости пикселя подаваемой на вход картинки.

Амплитуда импульсов напряжения от нейронов равна 0,5 В. Формирователь сигнала на периферии запоминающей матрицы на основе импульса от нейрона генерирует дополнительный импульс противоположной полярности (-0,5 В), который поступает на второй вход ячейки. Соответственно в зависимости от состояния мемристоров ячейки напряжение на выходе резистивного делителя может меняться от -0,5 до 0,5 В. При работе нейросети, потенциал на катоде диода равен потенциалу земли, так как выходы всех ячеек матрицы подключены к преобразователю ток-напряжение на основе ОУ. Диод является нагрузкой резистивного делителя и находится в закрытом состоянии.

Сумма выходных токов ячеек является входным синаптическим током для соответствующего нейрона. Электрическая схема выходного нейрона преобразует входной ток в напряжение, которое затем подается в интегрирующую схему с конденсатором. Превышение напряжения на конденсаторе выше заданного порога вызывает разряд конденсатора в

интеграторе и появление на выходе нейрона импульса напряжения с амплитудой 0,5 В и шириной 1 мс. Во время наличия выходного импульса нейрон игнорирует входной синаптический ток.

Для получения значений сопротивлений мемристоров кроссбара предварительно была обучена с учителем программная неимпульсная модель нейросети:

$$y_j = \frac{1}{1 + \exp(-s_j)}; s_j = w_{ij}x_i, \quad i \in [0; 48], \quad j \in [0; 9]$$

Входной вектор x_i был представлен значениями яркости пикселей. Порядок отображения двумерного изображения в одномерный вектор – слева направо, сверху вниз. В качестве активационной функции выходных программных нейронов использовалась сигмоида. Начальные значения матрицы весовых коэффициентов заданы случайно их равномерного распределения в интервале (-1; 1).

Режим обучения с учителем эффективен при наличии предварительно размеченного набора тестовых примеров, каковым и является выбранный датасет. Программный персептрон был обучен на основе дельта-правила изменения весов с коэффициентом 0,01.

$$\Delta w_{ij} = 0,01 \cdot (d_j - y_j) \cdot x_i$$

Вектор d_j соответствует ожидаемому выходному вектору. Всего было произведено сто эпох обучения по 50 000 обучающих примеров в каждой. Кривая обучения программной нейросети [85] представлена на рис.26.

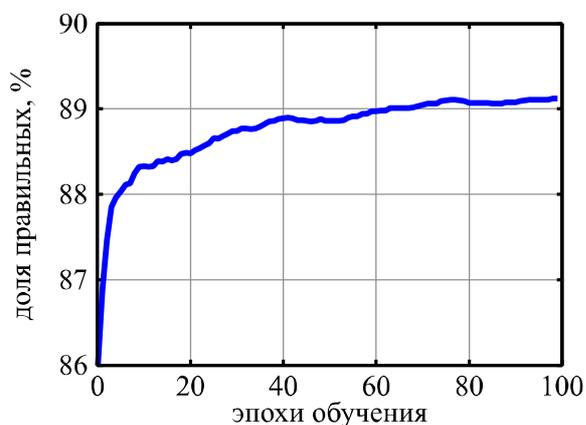


Рис.26. Кривая обучения программного персептрона с учителем.

Отображение получившихся значений весовых коэффициентов в сопротивлении мемристоров запоминающей матрицы выполнено с помощью линейного преобразования.

$$R_{ij}^+ = \frac{R_S}{2} (1 - k \cdot w_{ij}), \quad k = \frac{1}{\max(w_{ij})}, \quad R_S = R_{ij}^+ + R_{ij}^- = \text{const}$$

Здесь R_{ij}^+ и R_{ij}^- соответствуют сопротивлениям мемристоров M1 и M2 в ячейке, подключенной к i -тому входу и j -тому выходу запоминающей матрицы. Максимальное и минимальное сопротивления мемристоров равны 100 кОм и 1 кОм соответственно. Тогда $R_S = 101$ кОм.

Эквивалентность весовых коэффициентов для неимпульсной и импульсной нейросети достигается при использовании средней частоты спайков для кодирования информации (в данном случае яркости пикселей).

Пример работы аппаратного однослойного импульсного персептрона [86] с предварительно запрограммированной матрицей весовых коэффициентов показан на рис.27.

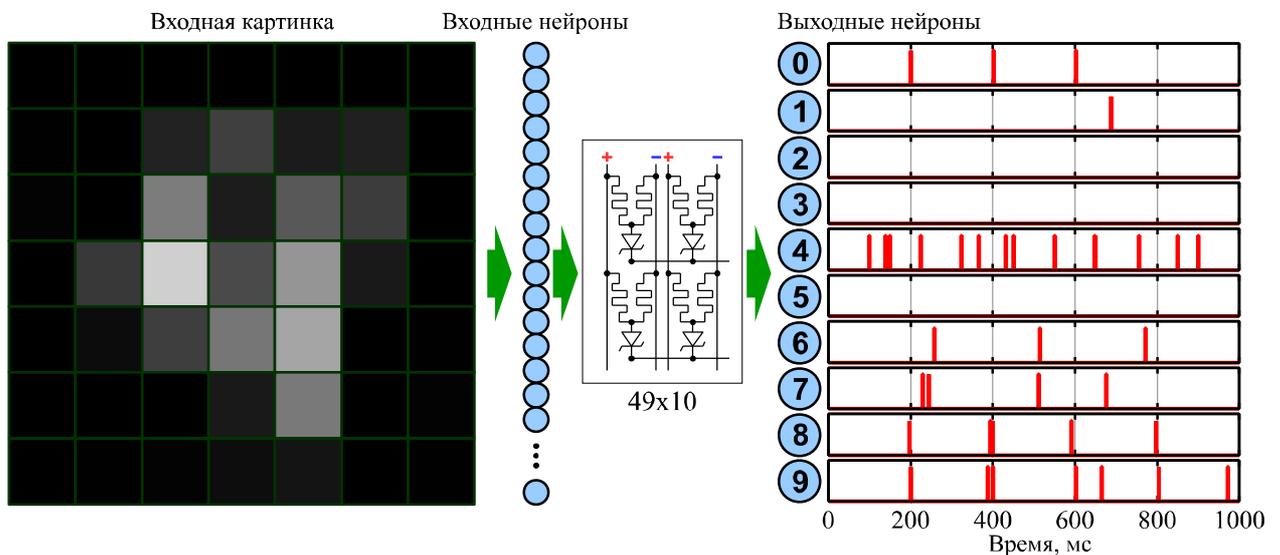


Рис.27. Результат распознавания цифры «4» из датасета MNIST импульсным персептроном с синапсами на основе мемристормо-диодного кроссбара.

В приведенном примере входная картинка соответствует рукописной цифре «4». В результате моделирования работы аппаратной нейросети в течении одной секунды видно, что нейросеть распознала входное

изображение как цифру «4», поскольку частота срабатывания выходного нейрона, ассоциированного с цифрой «4» является наибольшей (14 Гц).

4.2. Физико-математическая модель ассоциативного самообучения синапсов в запоминающей матрице на основе локального правила обучения LTP.

Самообучение синапсов в запоминающей матрице происходит по правилу Хебба, как и в реальном синапсе: сила связи между одновременно активировавшимися нейронами увеличивается. Изменение веса синапса зависит от разницы между временами срабатывания пресинаптического и постсинаптического нейронов $\Delta t = t_{pre} - t_{post}$.

С целью реализации известных правил индуцированной долговременной потенциации (long-term potentiation - LTP) и пластичности, зависимой от времени импульсов (spike-time dependent plasticity – STDP), разработаны физико-математические модели ассоциативного самообучения LTP и STDP в запоминающей матрице [87, 88, 92].

Оригинальная электрическая схема ячейки требует специфичной реализации правил самообучения. В обеих моделях самообучения пресинаптический сигнал представляет собой пару импульсов напряжения одинаковой амплитуды, но разной полярности вследствие использования комплементарной пары мемристоров в ячейке. Нейрон при срабатывании отправляет обратно в синапс импульс отрицательного напряжения *post*. Изменение силы синапса при этом произойдет если на синапс придут импульсы от пресинаптического нейрона. Форма пресинаптических импульсов определяет реализуемое локальное правило обучения.

Правило LTP предполагает увеличение веса синапса при совпадении во времени пресинаптических и постсинаптических сигналов. Поэтому в соответствующей модели знак импульсов напряжения пресинаптических сигналов поддерживается постоянным, что вызывает изменение состояние

мермисторов в ячейке всегда в сторону уменьшения сопротивления и увеличению веса синапса.

Весовой коэффициент синапса – ячейки запоминающей матрицы равен:

$$w = \frac{1}{R_d} \left(\frac{R_- - R_+}{R_- + R_+} \right),$$

где R_d - сопротивление диода Зенера, R_+ и R_- - сопротивления мемристоров. Согласно используемой SPICE модели мемристора (1) – (2) изменение его сопротивления определяется как:

$$\frac{dR}{dt} = f(V_R) \cdot W(R, V_R);$$

$$f(V_R) = \begin{cases} 0, & V_R \in [-V_t, V_t], \\ \beta \cdot V_R, & V_R \in (-\infty, -V_t) \cup (V_t, \infty); \end{cases}$$

$$W(R, V_R) = \begin{cases} 0, & V_R = 0 \vee R = R_{on} \vee R = R_{off}, \\ 1, & V_R \neq 0 \wedge (R > R_{on} \vee R < R_{off}). \end{cases}$$

Здесь V_R - напряжение на мемристоре.

Реализация правила LTP предполагает использования импульсов напряжения, форма которых приведена на рис.28а.

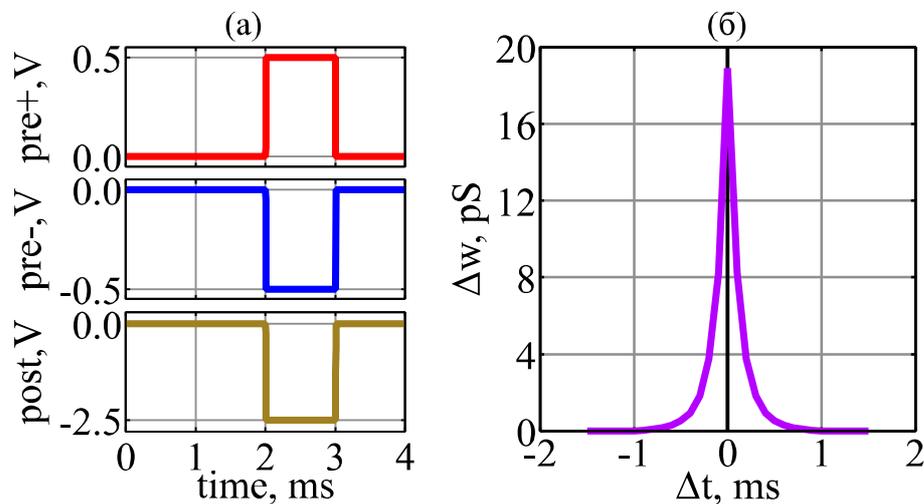


Рис.28. Долговременная потенцияция: (а) – пре и постсинаптические напряжения ячейки запоминающей матрицы при $\Delta t = 0$, (б) – функция пластичности $\Delta w(\Delta t)$.

Если пренебречь сопротивлением открытого диода Зенера, то напряжение $V_R \approx V_{pre} - V_{post}$. Напряжение на мемристоре, подключенном к

отрицательному входу, $V_R^- \approx V_{pre-} - V_{post}$ всегда меньше порога переключения, поэтому изменение его сопротивления не происходит.

Сопротивление мемристора, подключенного к положительному входу, будет изменяться при перекрытии пре- и постсинаптических импульсов напряжения:

При $\Delta t < 0$:

$$\Delta R_+(\Delta t) = \int_{t_{post}}^{t_{pre}+\tau} f(V_R^+)W(R_+, V_R^+) dt = \int_{t_{post}}^{t_{post}+\Delta t+\tau} \beta V_p dt = \beta V_R^+(\Delta t + \tau),$$

где t_{pre} и t_{post} - моменты времени переднего фронта пресинаптического и постсинаптического импульсов соответственно, τ - ширина импульсов, $V_p > V_t$ - напряжение программирования.

В случае $\Delta t > 0$:

$$\Delta R_+(\Delta t) = \int_{t_{pre}}^{t_{post}+\tau} f(V_R^+)W(R_+, V_R^+) dt = \int_{t_{pre}}^{t_{pre}-\Delta t+\tau} \beta V_p dt = \beta V_R^+(\tau - \Delta t).$$

Таким образом, с учетом $\Delta R_+(0) = 0$ и $\Delta R_-(\Delta t) = 0$ функция пластичности при изначальных состояниях мемристоров из середины возможного диапазона $R_+ = R_- = 0,5 \cdot (R_{off} + R_{on}) = R_0$ описывается формулой:

$$\Delta w(\Delta t) = \left[\frac{2R_0 R_d}{\beta V_p (\tau - |\Delta t|)} - R_d \right]^{-1}.$$

Вид функции пластичности при использовании миллисекундных импульсов представлен на рис.28б.

Аналитическое представление функции пластичности позволяет выбирать параметры в модели мемристора для достижения оптимальной скорости обучения нейросети, обеспечивающую сходимость численного расчета нейросети за минимальное время.

4.3. Физико-математическая модель ассоциативного самообучения синапсов в запоминающей матрице на основе локального правила обучения STDP.

Правило STDP действует аналогично LTP при $\Delta t < 0$. При $\Delta t > 0$, согласно этому правилу, вес синапса должен уменьшаться. В соответствующей модели такое уменьшение реализовано с помощью знакопеременных импульсов напряжения (рис.29а). Итоговое направление изменения состояния мемристоров и веса синапсов определяется преобладающими знаками напряжения пресинаптических импульсов во время действия постсинаптического импульса.

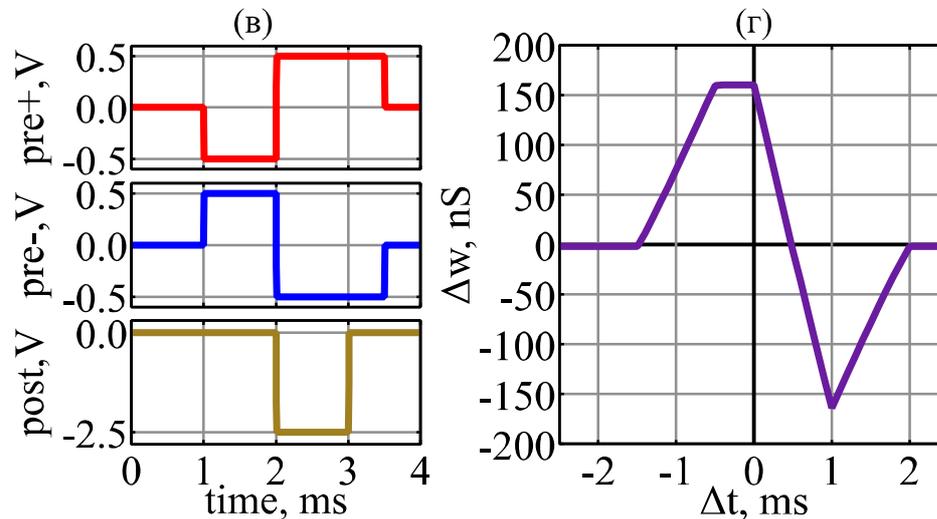


Рис.29. Пластичность, зависящая от времени импульса: (а) – пре и постсинаптические напряжения ячейки запоминающей матрицы при $\Delta t = 0$, (б) – функция пластичности $\Delta w(\Delta t)$.

Ширина начальной фазы пресинаптического импульса и постсинаптического импульса равна τ , ширина второй фазы пресинаптического импульса равна $\tau_1 > \tau$. Из-за сложной формы импульсов в синапсе функция пластичности является кусочной и состоит из 6 кривых:

$$1) \Delta w(\Delta t) = 0, \text{ при } \Delta t > 2\tau;$$

$$2) \Delta w(\Delta t) = \left[R_d - \frac{2R_0R_d}{\beta V_p(2\tau - \Delta t)} \right]^{-1}, \text{ при } \tau < \Delta t < 2\tau;$$

$$3) \Delta w(\Delta t) = \left[\frac{2R_0 R_d}{\beta V_p \tau} - R_d + 2R_d \frac{\Delta t}{\tau} \right]^{-1}, \text{ при } 0 < \Delta t < \tau;$$

$$4) \Delta w(\Delta t) = \left[\frac{2R_0 R_d}{\beta V_p \tau} - R_d \right]^{-1} = \text{const}, \text{ при } \tau - \tau_1 < \Delta t < 0;$$

$$5) \Delta w(\Delta t) = \left[\frac{2R_0 R_d}{\beta V_p (\tau_1 + \Delta t)} - R_d \right]^{-1}, \text{ при } -\tau_1 < \Delta t < \tau - \tau_1;$$

$$6) \Delta w(\Delta t) = 0, \text{ при } \Delta t < -\tau_1.$$

Вид функции пластичности показан на рис.29б. Преобладание области с усилением синапса связано с $\tau_1 > \tau$, а смещение точки с нулевым изменением веса – с выбором момента отчета времени пресинаптического импульса t_{pre} .

4.4. Численное моделирование процесса ассоциативного самообучения запоминающей матрицы в составе аппаратной импульсной нейросети с использованием экспериментальных ВАХ ячеек.

При моделировании долговременной потенциации (LTP) [1] выходные импульсы двух нейронов через синаптические связи, представленные комплементарными мемристорно-диодными ячейками, поступают на вход третьего нейрона (рис.30).

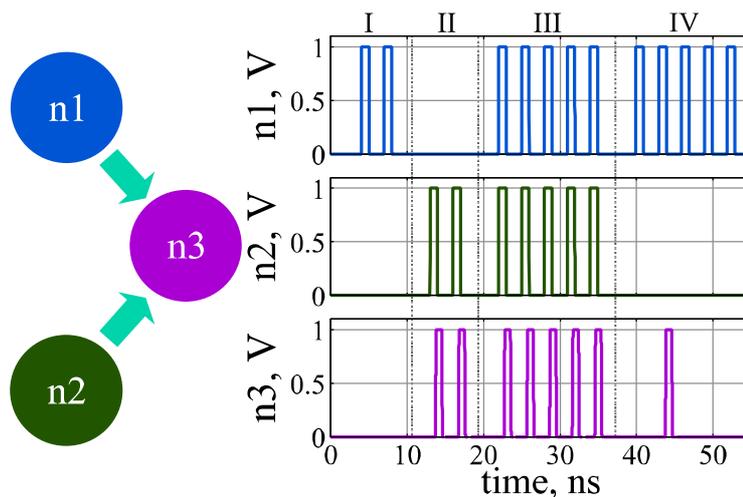


Рис.30. Возникновение новой ассоциации в группе из трех нейронов.

Импульсы напряжения, генерируемые нейронами на разных этапах ассоциативного обучения.

Моделирование этой же нейросети (рис.30) с пластичностью, зависимой от времени импульса (STDP), приводит к тому же результату, поскольку ситуации с $\Delta t > 0$ отсутствуют.

Проведено численное моделирование ассоциативного самообучения аппаратной импульсной нейросети в виде однослойного персептрона с использованием запоминающей матрицы в качестве массива синапсов [89]. Электрические параметры ячеек при моделировании соответствовали усредненным характеристикам мемристорно-диодного кроссбара (рис.31) с числом ячеек 4×2 , который изготовлен с помощью магнетронной технологии [84]. Ячейки матрицы 1D2M представляют собой двухслойное соединение комплементарных биполярных мемристоров и одного разделяющего диода Зенера. Изготовление мемристорно-диодного кроссбара $\text{TiN}/\text{Ti}_{0,93}\text{Al}_{0,07}\text{O}_x/\text{p-Si}/\text{n-Si}/\text{W}$ (рис. 31а) проводили в магнетронном модуле, входящем в нанотехнологический комплекс «НаноФаб-100» (NT-MDT, Россия).

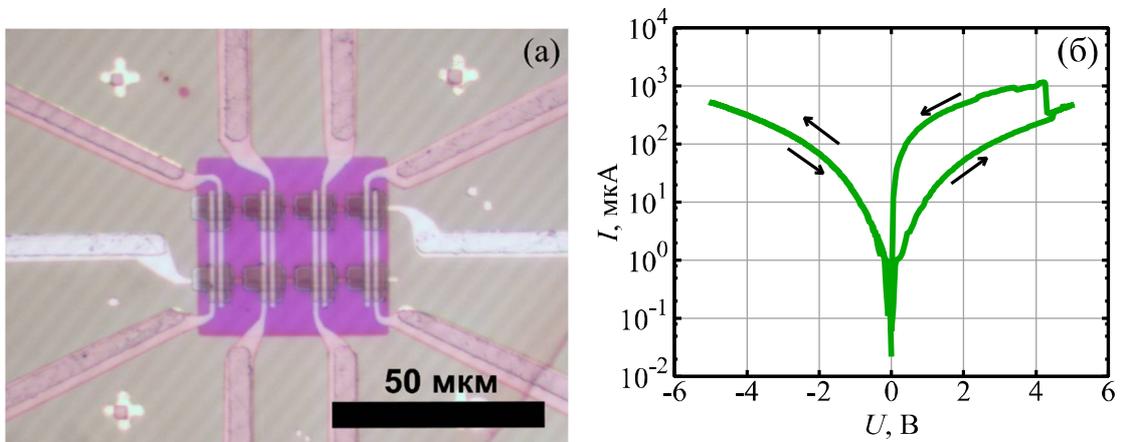


Рис.31. Микрофотография изготовленного мемристорно-диодного кроссбара $\text{TiN}/\text{Ti}_{0,93}\text{Al}_{0,07}\text{O}_x/\text{p-Si}/\text{n-Si}/\text{W}$ - (а); усредненная ВАХ его ячеек - (б).

Слой смешанного оксида металлов осаждали при одновременном распылении двух катодов из титана (Ti) и алюминия (Al), а полупроводниковые слои диода – при последовательном распылении катодов из легированного бором и фосфором кремния (p-Si и n-Si соответственно). Мемристоры изготавливали по технологии кроссбар путем последовательного напыления пленок через маски электронного резиста (полиметилметакрилат),

выполненные на электронном микроскопе JSM-6510LV-EDS (Jeol, Япония). Вольт-амперная характеристика (ВАХ) ячейки кроссбара приведена на рис.31б.

Максимальное отношение сопротивлений ячейки кроссбара в низко- и высокопроводящем состояниях достигается при напряжении $U = 0.3$ В и составляет 54.2. Среднее сопротивление ячеек в низкопроводящем состоянии равно 1.56 МОм со среднеквадратичным отклонением 0.43 МОм при $U = 2$ В. Характеристики изготовленных мемристоров в составе мемристорно-диодного кроссбара исследовали ранее в работах [90, 91]. Удельные сопротивления предельных состояний при циклических переключениях ρ_{off} и ρ_{on} составляют 1.14 ± 0.18 кОм·м и 1.61 ± 0.05 Ом·м соответственно. Напряжение переключения ячейки оказалось больше, чем у отдельного мемристора [84], поскольку часть напряжения падает на диоде. Большое сопротивление закрытого диода приводит к стягиванию гистерезиса в обратной ветви ВАХ ячейки, так как вклад сопротивления диода преобладает над вкладом малого сопротивления мемристора в их общую ВАХ. Снижение напряжения переключения ячейки возможно при уменьшении толщины активного слоя мемристора и увеличении нелинейности селективного элемента. Распределенный характер нейросети снижает требования к воспроизводимости и стабильности характеристик мемристоров, и на первый план выходит отношение предельных значений сопротивлений для существования множества синаптических состояний. Это подтверждается в работе [84], где работоспособность кроссбара сохранялась при разбросе предельных сопротивлений ячеек до 100%.

Электрическая схема аппаратной импульсной нейросети (рис.32) построена на основе мемристорно-диодного кроссбара (рис.31) с 4-мя парами входных проводников и двумя выходными шинами [88, 89]. Соответственно, кроссбар содержит восемь ячеек, которые являются синапсами нейросети. Слой персептрона образован двумя нейронами, построенными на основе операционных усилителей.

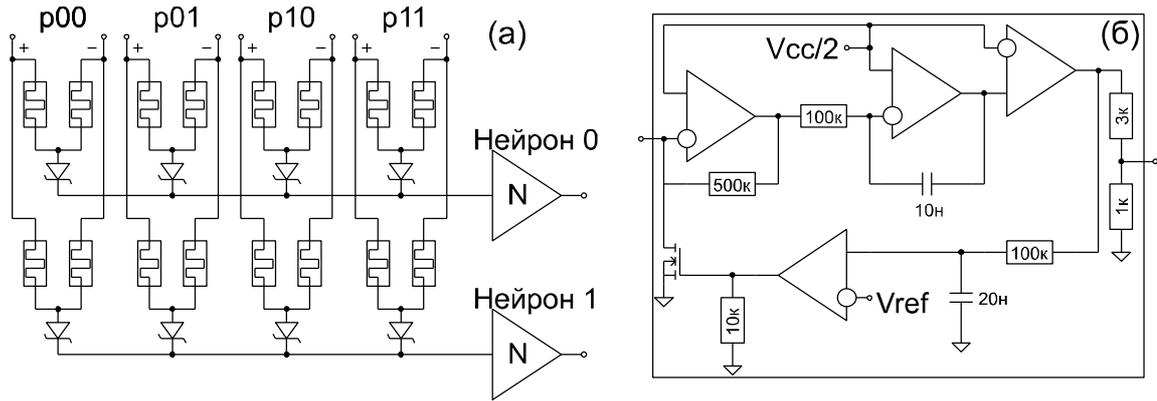


Рис.32. Электрическая схема аппаратной нейросети:

- а) мемристорно-диодный кроссбар в качестве массива синапсов,
- б) электрическая схема нейрона.

Электрическая схема нейрона состоит из преобразователя ток–напряжение, аналогового интегратора, компаратора, схемы задержки в виде интегрирующей RC-цепи и полевого транзистора.

Нейросеть, состоящая из четырех виртуальных входных нейронов и двух аппаратных выходных нейронов (рис.33) [89], обучалась по правилу STDP для выполнения задачи распознавания картинок с разрешением 2x2 пикселя.

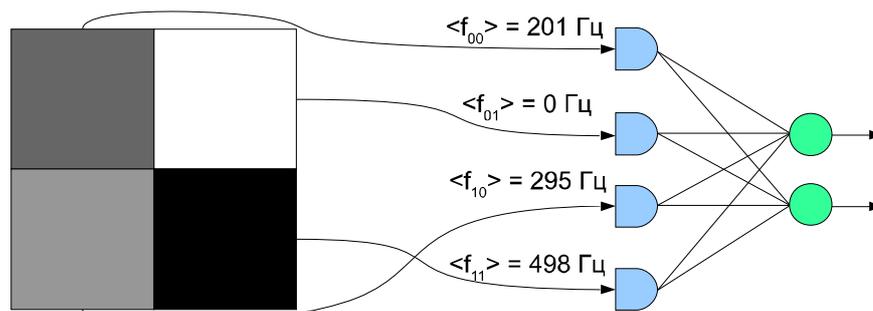


Рис.33. Входное изображение и архитектура аппаратной нейросети: синий цвет – входные виртуальные нейроны, зеленый – выходные нейроны.

Значения яркостей пикселей преобразовывались в среднюю частоту последовательности входных импульсов с помощью микроконтроллера. Выходные импульсы активировавшихся нейронов фиксировались этим же микроконтроллером. Результат распознавания изображения определял

выходной нейрон сети с большим числом выходных импульсов, зарегистрированных во время присутствия последовательности входных импульсов. Каждый цикл обучения (эпоха) состоит из последовательной подачи на вход нейросети импульсов, соответствующих 128 изображениям, образованных путем добавления шума к исходным двум эталонным картинкам. В каждой эпохе набор изображений одинаковый.

Модельная кривая обучения, представленная на рис.34а, показывает долю правильно классифицированных входных картинок от числа циклов (эпох) обучения [88].

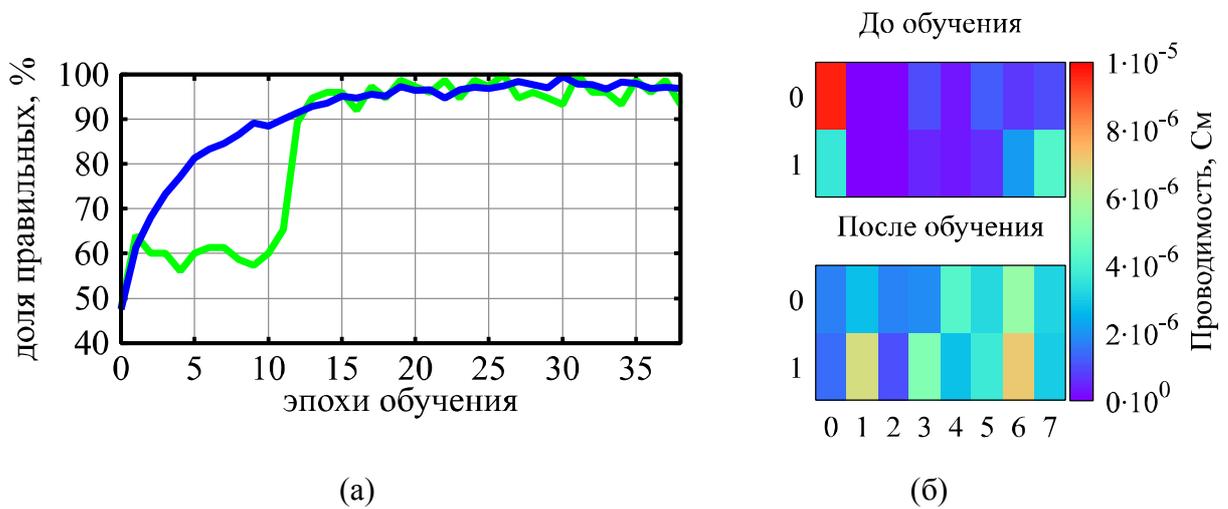


Рис.34. Результат ассоциативного самообучения аппаратной нейросети:

- а) изменение доли правильных классификаций: модельная (синяя) и экспериментальная (зеленая) кривые; б) изменение проводимости мемристоров кроссбара.

Для проверки адекватности физико-математической модели ассоциативного самообучения STDP на рис.34 кроме модельной кривой изменения доли правильных классификаций картинок показана аналогичная экспериментальная кривая, полученная при обучении изготовленной аппаратной нейросети (рис.32) [88]. Модельная кривая обучения более гладкая по сравнению с экспериментальной. Это объясняется тем, что изменение проводимости в модели мемристора происходит плавно, без скачков. Расположение прямоугольников на рис.34б соответствует

положению мемристоров на рис.32а. Изначально проводимость мемристоров была низкой за исключением одного мемристора. В процессе ассоциативного самообучения проводимость этого мемристора уменьшилась, а остальных выросла.

На рис.35 показан суммарные входные импульсы напряжения нейрона после преобразователя ток-напряжение, полученные в результате SPICE моделирования процесса ассоциативного самообучения аппаратной импульсной нейросети с мемристорно-диодным кроссбаром [88].

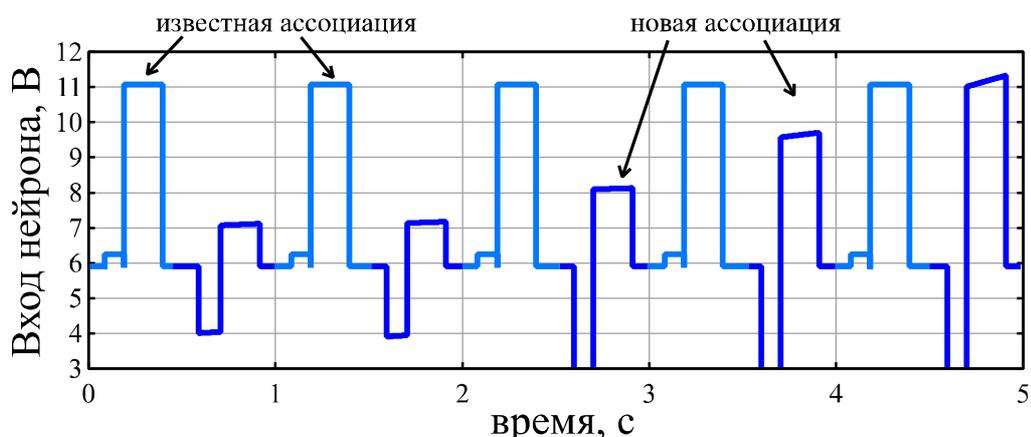


Рис.35. Результат SPICE моделирования процесса генерации новой ассоциации на фоне известной.

Сначала подаются пресинаптические импульсы напряжения на оба синапса и происходит потенциация слабой связи между нейронами. Соответствующий входной сигнал для нейрона показан светло-синим цветом. Затем импульс подается только на слабый синапс, что порождает слабый входной импульс напряжения для нейрона (темно-синий цвет). Потенциация слабого синапса проявляется в увеличении амплитуды второго импульса и происходит из-за того, что они активируются одновременно, и напряжение на мемристорах, соответствующих этой связи, выше порогового.

Для демонстрации процесса быстрого формирования новой ассоциации были выбраны импульсы большой длительности, чтобы переобучение произошло за небольшое количество импульсов. Этот же процесс описан в

[9], где представлен начальный и конечный результат формирования новой ассоциации после большого числа импульсов.

Нарастание входного напряжения нейрона, которое пропорционально синаптическому току, является признаком формирования новой ассоциации и вызвано усилением синапса согласно правилу STDP. Генерация новой ассоциации при переобучении происходит в изготовленном мемристорно-диодном кроссбаре в отличие от существующих нейросетей с синапсами на базе дискретных мемристоров.

4.5. Выводы к главе 4

Показан способ отображения входного вектора слоя нейросети и её весовых коэффициентов на входной вектор напряжений и сопротивления мемристоров соответственно.

В качестве проверки возможности использования разработанной запоминающей матрицы в качестве массива синапсов аппаратной импульсной нейросети построена сеть прямого распространения информации – однослойный аппаратный персептрон. Для получения значений сопротивлений мемристоров кроссбара предварительно была обучена с учителем программная неимпульсная модель нейросети. Моделирование аппаратной импульсной нейросети покало успешное распознавание цифр из датасета MNIST, предварительно уменьшенных в разрешении до 7x7 пикселей.

С целью реализации известных правил индуцированной долговременной потенциации (long-term potentiation - LTP) и пластичности, зависимой от времени импульсов (spike-time dependent plasticity – STDP), разработаны физико-математические модели ассоциативного самообучения LTP и STDP в запоминающей матрице.

Проведены численное моделирование и экспериментальное исследование по ассоциативному самообучению аппаратной импульсной нейросети по правилу STDP в виде однослойного персептрона с

использованием запоминающей матрицы в качестве массива синапсов. Для обеспечения более высокой точности физико-математических моделей самообучения нейросети и совпадения результатов расчета с экспериментальными результатами электрические параметры ячеек при моделировании соответствовали усредненным характеристикам ячеек мемристорно-диодного кроссбара размером 4×2 , который был изготовлен с помощью магнетронной технологии.

Многослойная структура мемристорно-диодного кроссбара изготовлена методом магнетронного распыления в одном технологическом модуле. Такой подход позволил упростить существующий технологический цикл производства, в котором изготавливаются: активный слой мемристора методом атомно-слоевого осаждения, проводники методом магнетронного распыления, полупроводниковые слои диода методами термодиффузии и ионной имплантации.

В ходе численного моделирования явно показан процесс генерации новой ассоциации в аппаратной импульсной нейросети по правилу STDP с массивом синапсов в виде разработанной запоминающей матрицы на основе мемристорно-диодного кроссбара.

ЗАКЛЮЧЕНИЕ

1. Созданы физические основы, электрическая схема и физико-математическая модель работы ячейки сверхбольшой запоминающей матрицы на основе комплементарно включенных мемристоров и диода Зенера. Комплементарное включение двух мемристоров обеспечивает малое энергопотребление. Выбор в качестве селективного элемента диода Зенера, обеспечивает существенно большую интеграцию элементов и меньшее энергопотребление по сравнению с активным селектором – транзистором.

2. Разработаны электрическая схема и физико-математическая модель процессов обработки сигналов в сверхбольшой запоминающей матрице, представляющей собой массив синаптических связей аппаратной нейросети, реализующейся в нейропроцессоре на основе мемристорно-диодного кроссбара и обладающей высокими интеграцией элементов и энергоэффективностью.

3. С помощью численного моделирования показано, что применение диода Зенера в качестве селективного элемента, подключенного электрически последовательно мемристоры, позволяет увеличить размер мемристорно-диодного кроссбара запоминающей матрицы вплоть до 10^6 ячеек за счет уменьшения паразитных токов между ячейками, а также уменьшить деградацию выходного сигнала при суммировании входных импульсов напряжения в матрице. Показано, что разработанная запоминающая матрица обладает меньшей деградацией сигнала и большей энергоэффективностью по сравнению с известными аналогами.

4. Предложена топология сверхбольшой запоминающей матрицы на кристалле в планарном исполнении и с применением 3D интеграции. В результате сравнения с известными аналогами показано, что разработанная запоминающая матрица обладает высокой интеграцией элементов.

5. С помощью численного моделирования показана работоспособность малой матрицы размером 2×2 ячеек при взвешивании импульсов напряжения и сложении токов ячеек. Результаты численного

моделирования хорошо согласуются с экспериментальными данными, что подтверждает корректность разработанной физико-математической модели работы матрицы и дает основание для использования этой модели при проектировании матриц сверхбольшого размера.

6. Разработаны физико-математические модели ассоциативного самообучения запоминающей матрицы, которые обеспечивают специфичную реализацию правил самообучения LTP и STDP, связанную с оригинальностью мемристорно-диодной ячейки.

7. При численном моделировании были использованы экспериментальные характеристики ячеек запоминающей матрицы для обеспечения более высокой точности физико-математических моделей самообучения нейросети. С этой целью был изготовлен многослойная структура мемристорно-диодного кроссбара методом магнетронного распыления в одном технологическом модуле.

8. С помощью численного моделирования продемонстрирована способность запоминающей матрицы с числом ячеек 4×2 к ассоциативному самообучению в составе аппаратной импульсной нейросети. Результаты численного моделирования хорошо согласуются с экспериментальными данными, что подтверждает корректность разработанной физико-математической модели обучения по правилу STDP и дает основание для использования этой модели при создании матрицы сверхбольшого размера.

9. Продемонстрирована генерация новой ассоциации (нового знания) при численном моделировании аппаратной импульсной нейросети на основе мемристорно-диодного кроссбара в отличие от ассоциативного самообучения в существующих аппаратных нейросетях с синапсами на базе дискретных мемристоров.

СПИСОК ЛИТЕРАТУРЫ

1. Pisarev A.D., Busygin A.N., Udovichenko S.Yu., Maevsky O.V. A biomorphic neuroprocessor based on a composite memristor-diode crossbar // *Microelectronics Journal*. 2020. V. 102. Article 104827.
2. Filippov V.A., Bobylev A.N., Busygin A.N., Pisarev A.D., Udovichenko S.Yu. A biomorphic neuron model and principles of designing a neural network with memristor synapses for a biomorphic neuroprocessor // *Neural Computing and Applications*. 2020. V.32. P. 2471-2485.
3. Li C., Belkin D., Li Y. et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks // *Nature Communications*. 2018. V. 9. Article 2385.
4. Ielmini D. Brain-inspired computing with resistive switching memory (RRAM): Devices, synapses and neural networks // *Microelectronic Engineering*. 2018. V. 190. P. 44–53.
5. Yao P., Wu H., Gao B. et al. Fully hardware-implemented memristor convolutional neural network // *Nature*. 2020. V. 577. P. 641-661.
6. Cai F., Correl J.M., Lee S.H. et al. A fully integrated reprogrammable memristor–CMOS system for efficient multiply–accumulate operations // *Nature Electronics*. 2019. V. 2. P. 290–299.
7. Wang Z., Li C., Song W. et al. Reinforcement learning with analogue memristor arrays // *Nature Electronics*. 2019. V. 2. P. 115–124.
8. Wang Z., Wang X., A Novel Memristor-Based Circuit Implementation of Full-Function Pavlov Associative Memory Accorded with Biological Feature // *IEEE Transactions on Circuits and Systems I:Regular Papers*. 2018 .V. 65(7). P. 2210-2220.
9. Minnekhanov A.A., Emelyanov A.V., Lapkin D.A. et al. Parylene Based Memristive Devices with Multilevel Resistive Switching for Neuromorphic Applications // *Sci. Rep*. 2019. V. 9. P. 10800.

10. Demin V.A. Nekhaev D.V., Surazhevsky I.A. et al. Necessary conditions for STDP-based pattern recognition learning in a memristive spiking neural network // *Neural Networks*. 2021. V. 134. P. 64-75.
11. Antonov I. N., Belov A. I., Mikhaylov A. N. et al. Formation of Weighting Coefficients in an Artificial Neural Network Based on the Memristive Effect in Metal–Oxide–Metal Nanostructures // *Journal of Communications Technology and Electronics*. 2018. V. 63(8). P. 950–957.
12. Surazhevsky I.A., Demin V.A., Ilyasov A.I. et al. Noise-assisted persistence and recovery of memory state in a memristive spiking neuromorphic network // *Chaos, Solitons and Fractals*. 2021. V. 146. Article 110890
13. Schuman C.D., Potok T.E., Patton R.M. et al. A Survey of Neuromorphic Computing and Neural Networks in Hardware // *arXiv preprint*. 2017. Article 1705.06963.
14. Silberberg G., Gupta A., Markram H. Stereotypy in neocortical microcircuits // *Trends in Neurosciences*. 2002. V. 25(5). P. 227-230.
15. Lin A., Rivenson Y., Yardimei N.T. et al. All-optical machine learning using diffractive deep neural networks // *Science*. 2018. V. 361. P.1004-1008.
16. Ionica M.H., Gregg D. The Movidius Myriad Architecture's Potential for Scientific Computing // *IEEE Micro*. 2015. V. 35(1). P. 6-14.
17. Jouppi N., Young C., Patil N., Patterson D. Motivation for and Evaluation of the First Tensor Processing Unit // *IEEE Micro*. 2018. V. 38(3). P. 10-19.
18. Shafiee A., Nag A., Muralimanohar N. et al. ISAAC: A Convolutional Neural Network Accelerator with In-Situ Analog Arithmetic in Crossbars // *2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, 2016, p. 14-26.
19. Ankit A., Hajj I.E., Chalamalasetti S.R. et al. PUMA: A Programmable Ultra-efficient Memristor-based Accelerator for Machine Learning Inference // *2019 Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2019, p. 715-731.

20. https://mipt.ru/news/laboratoriya_neyrovychislitelnykh_sistem_mfti_sozdala_pervyy_otechestvennyy_neyrosetevoy_protsessor
21. Akida Neuromorphic System-on-Chip [Online]. Available: <https://brainchipinc.com/akida-neural-processor-soc/>
22. Merolla P.A., Arthur J.V., Alvarez-Icaza R., et al. A million spiking-neuron integrated circuit with a scalable communication network and interface // Science. 2014. V.345. P.668-672.
23. Van Der Made P. A.J., Viejo A., Mankar A.S., Viejo M. Neural Processor based accelerator system and method // US Patent № 2017/0024644 A1, 2017.
24. Davies M., Srinivasa N., Lin T.-H., et al. Loihi: A Neuromorphic Manycore Processor with On-Chip Learning // IEEE Micro. 2018. V.38(1). P.82-99.
25. Prezioso M., Merrih-Bayat F., Hoskins B.D. et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors // Nature. 2015. V. 521. P. 61-64.
26. Liu T., Yan T.H., Scheuerlein R. et al. 130.7mm² 2-Layer 32-Gb ReRAM Memory Device in 24-nm Technology // IEEE Journal of Solid-State Circuits. 2014. V. 49(1). P. 140-153.
27. Kim, K.-H., Gaba S., Wheeler D. et al. A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications // Nano Lett. 2012. V. 12. P. 389–395.
28. Jo S. H., Chang T., Ebong I. et al. Nanoscale Memristor Device as Synapse in Neuromorphic Systems // Nano Lett. 2010. V. 10. P.1297-1301.
29. Удовиченко С.Ю., Бобылев А.Н., Бусыгин А.Н., Писарев А.Д., Филиппов В.А. Прототип нейроморфного сопроцессора на основе мемристоров из смешанного оксида металлов // Сборник тезисов VI конференции Нанотехнологического общества России, 2016, с.29-32.
30. Bobilev A.N., Udovichenko S.Yu. The electrical properties of memristor devices TiN/Ti_x Al_{1-x} O_y/TiN produced by magnetron sputtering // Russian Microelectronics. 2016. V.45(6). P.396-401.

31. Pershin Y.V., Di Ventra M. Experimental demonstration of associative memory with memristive neural networks // *Neural Networks*. 2010. V. 23. № 7. P.881-886.
32. Zhao W., Portal J., Kang W. et al. Design and analysis of crossbar architecture based on complementary resistive switching non-volatile memory cells // *Journal of Parallel and Distributed Computing*. 2014. V. 74(6). P. 2484-2496.
33. Chabi D., Querlioz D., Zhao W., Klein J.-O. Robust Learning Approach for Neuro-Inspired Nanoscale Crossbar Architecture // *ACM Journal on Emerging Technologies in Computing Systems (JETC)*. 2014. V. 10(1). Article 5.
34. Chabi D., Zhaohao W., Bennet C. et al. Ultra High Density Memristor Neural Crossbar for On-chip Supervised Learning // *IEEE Transactions on Nanotechnology*. 2015. V. 14(6). P.954-962.
35. Chabi D., Querlioz D., Zhao W., Klein J.-O. On-Chip Universal Supervised Learning Methods for Neuro-Inspired Block of Memristive Nanodevices // *ACM Journal on Emerging Technologies in Computing Systems (JETC)*. 2015. V. 11(4). Article 34.
36. Liu J., Harkin J, Maguire L.P. et al. SPANNER: A Self-Repairing Spiking Neural Network Hardware Architecture // *IEEE Transactions on Neural Networks and Learning Systems*. 2017. V. 29(4). P.1287-1300.
37. Chevallier C.J., Siau C.H., Lim S.F. et al. A 0.13 μm 64Mb Multi-Layered Conductive Metal-Oxide Memory // *Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2010, p. 260-261.
38. Bandyopadhyay A., Scheuerlein R.E., Gorla C.R., Le B. FET low current 3D ReRAM non-volatile storage // *US Patent № 0070965 A1.12.03.2015*.
39. Bennet C., Querlioz D., Klein J.-O. Spatio-temporal Learning with Arrays of Analog Nanosynapses // *2017 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, 2017, p. 125-130.
40. Li C., Hu M., Li Y. et al Analogue signal and image processing with large memristor crossbars // *Nature electronics*. 2018. V. 1(1). P. 52–59.

41. Ghenzi N., Rozenberg M., Pietrobon L. et al. One-transistor one-resistor (1T1R) cell for large-area electronics // *Applied Physics Letters*. 2018. V. 113. Article 072108.
42. 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. N. 9. Article 1705914.
43. Ambrogio S., Narayanan P., Tsai H. et al. Equivalent-accuracy accelerated neural-network training using analogue memory // *Nature*. 2018. V. 558. P. 60-67.
44. Shi Yu., Nguyen L., Oh S. et al. Neuroinspired unsupervised learning and pruning with subquantum CBRAM arrays // *Nature Communications*. 2018. V. 9. Article 5312.
45. Yao P., Wu H., Gao B. et al. Online Training on RRAM based Neuromorphic Network: Experimental Demonstration and Operation Scheme Optimization // *IEEE Electron Devices Technology and Manufacturing Conference (EDTM)*, 2017, p. 182-183.
46. Levisse A., Gaillardon P.E., Giraud B. et al. Resistive Switching Memory Architecture Based on Polarity Controllable Selectors // *IEEE Transactions On Nanotechnology*. 2019. V. 18. P. 183-194.
47. Zhang Y., Shen Y., Wang X., Cao L. A Novel Design for Memristor-Based Logic Switch and Crossbar Circuits // *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2015. V. 62(5). P. 1402-1411.
48. Teimoori M., Amirsoleimani A., Ahmadi A., Ahmadi M. A 2M1M Crossbar Architecture: Memory // *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 2018. V. 26(12). P. 2608-2618.
49. Huang J.-J., Tseng Y.-M., Luo W.-Ch. et al. One Selector-One Resistor (1S1R) Crossbar Array for High-density Flexible Memory Applications // *2011 International Electron Devices Meeting*, 2011, p. 31.7.1-31.7.4.
50. Zhang L., Govoreanu B., Redolfi A. et al. High-drive current ($>1\text{MA}/\text{cm}^2$) and highly nonlinear ($>10^3$) TiN/amorphous-Silicon/TiN scalable bidirectional selector with excellent reliability and its variability impact on the 1S1R array

performance // 2014 IEEE International Electron Devices Meeting, 2014, p. 6.8.1-6.8.4.

51. Choi B.J., Zhang J., Norris K. et al. Trilayer Tunnel Selectors for Memristor Memory Cells // *Advanced Materials*. 2016. V. 28. N. 2. P. 356-362.

52. Seo Sh., Lim J., Lee S. et al. Graphene-Edge Electrode on a Cu-Based Chalcogenide Selector for 3D Vertical Memristor Cells // *ACS Applied Materials & Interfaces*. 2019. V. 11. P.43466-43472.

53. Lee D.K., Kim G.H., Sohn H., Yang M.K. Positive effects of a Schottky-type diode on unidirectional resistive switching devices // *Applied Physics Letters*. 2019. V. 115. Article 263502

54. Yoon J.H., Kwon D.E., Kim Y. et al. The current limit and self-rectification functionalities in the TiO₂/HfO₂ resistive switching material system // *Nanoscale*. 2017. V. 9. P. 11920-11928.

55. Song M., Lee H., Seo D.H. et al. Improved Distribution of Resistance Switching Through Localized Ti-Doped NiO Layer With InZnO_x/CuO_x Oxide Diode // *IEEE Journal of the Electron Devices Society*. 2018. V. 6. P. 905-909.

56. Oh S.-I, Rani J.R., Hong S.-M., Jang J.-H. Self-rectifying bipolar resistive switching memory based on an iron oxide and graphene oxide hybrid // *Nanoscale*. 2017. V. 9. Article. 15314.

57. Luo Q., Zhang X., Hu Y. et al. Self-Rectifying and Forming-Free Resistive-Switching Device for Embedded Memory Application // *IEEE Electron Device Letters*. 2018. V. 39(5). P. 664-667.

58. Wong S., Hu C.M. SPICE macro model for the simulation of zener diode I-V characteristics // *IEEE Circuits and Devices Magazine*. 1991. V. 7(4). P. 9-12.

59. Bobylev A.N., Udovichenko S.Yu., Busygin A.N., Ebrahim A.H. Increase of switching range of resistive memristor for realization of a greater number of synaptic states in a neuroprocessor // *Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy*. 2019. V. 5(2). P. 124-136.

60. Vinet M., Batude P., Tabone C. et al. 3D monolithic integration: Technological challenges and electrical results // *Microelectronic Engineering*. 2011. V. 88(4). P. 331-335.
61. Shulaker M.M., Hills G, Park R.S. et al. Three-dimensional integration of nanotechnologies for computing and data storage on a single chip // *Nature*. 2017. V. 547. P. 74-78.
62. Lupan O., Pauporté Th., Tiginyanu I.M. et al. Optical properties of ZnO nanowire arrays electrodeposited on n- and p-type Si(1 1 1): Effects of thermal annealing // *Materials Science and Engineering: B*. 2011. V. 176(16). P. 1277-1284.
63. Abe H., Fujishima M., Komiyama T. et al. Heterojunction characteristics of ZnO and CuO substrates formed by direct bonding // *Phys. Status Solidi C*. 2012. V. 9(6). P. 1396-1399.
64. Nikiruy K.E., Emelyanov A.V., Rylkov V.V. et al. Adaptive Properties of Spiking Neuromorphic Networks with Synapses Based on Memristive Elements // *Technical Physics Letters*. 2019. V. 45(4). P. 386–390.
65. Demin V.A., Erokhin V.V., Emelyanov A.V. et al. Hardware elementary perceptron based on polyaniline memristive devices // *Organic Electronics*. 2015. V. 25. P.16–20.
66. Emelyanov A.V., Lapkin D.A., Demin V.A. et al. First steps towards the realization of a double layer perceptron based on organic memristive devices // *AIP Advances*. 2016. V. 6. Article 111301.
67. Xia O., Robinett W., Cumbie M. et al. Memristor-CMOS hybrid integrated circuits for configurable logic // *Nano Letters*. 2009. 9(10). P.3640-3645.
68. Maevsky O.V., Pisarev A.D., Busygin A.N., Udovichenko S.Yu. Complementary memristive diode cells for the memory matrix of a neuromorphic processor // *International Journal of Nanotechnology*. 2018. V.15. №4/5. P.388-393.

69. Pisarev A., Busygin A., Udovichenko S., Maevsky O. 3D memory matrix based on a composite memristor-diode crossbar for a neuromorphic processor // *Microelectronic Engineering*. 2018. V.198. P.1-7.
70. Kim H.K., Li C.C., Fang X.M. et al. Erbium Doped Semiconductor Thin Films Prepared by RF Magnetron Sputtering // *Materials Research Society Symposia Proceedings*. 1993. V. 301. P. 55-60.
71. Удовиченко С.Ю., Маевский О.В., Писарев А.Д., Бусыгин А.Н., Комплементарная мемристорно-диодная ячейка для запоминающей матрицы нейроморфного процессора. // *Сборник тезисов VIII конференции Нанотехнологического общества России*. 2017. С. 37-40.
72. Маевский О.В., Писарев А.Д., Бусыгин А.Н., Удовиченко С.Ю. Запоминающее устройство на основе комплементарной мемристорно-диодной ячейки // *Патент РФ №2649657*. Дата публикации патента 04.04.2018.
73. Keeth B., Baker R., Johnson B., Lin F. *DRAM Circuit Design: Fundamental and High-Speed Topics, Second Edition* // Wiley-IEEE Press. 2007. P. 440.
74. Baker. *R.CMOS: Circuit Design, Layout, and Simulation, 3rd Edition* // Wiley-IEEE Press. 2010. P. 1208.
75. Udovichenko S.Yu., Pisarev A.D., Busygin A.N., Bobylev A.N. Memory and universal logic matrices for neuroprocessor. // *First International Workshop "Nanoelectronic Memristive Devices for Quantum and Neuromorphic Computing"*. Book of Abstracts. 2018. P. 19.
76. Удовиченко С.Ю., Писарев А.Д., Бусыгин А.Н., Бобылев А.Н. Высокая интеграция элементов в логической и запоминающей матрицах нейропроцессора с помощью комбинированного мемристорно – диодного кроссбара // *Сборник тезисов IX конференции Нанотехнологического общества России*. 2018. С. 6-7.
77. Писарев А.Д., Бусыгин А.Н., Удовиченко С.Ю., Бобылев А.Н., Маевский О.В. 3D запоминающая матрица на основе комплементарной

мемристорно-диодной ячейки // Патент РФ №2697623. Дата публикации патента 15.08.2019.

78. Chen W., Lin X., Parris P.M. Zener diode device and fabrication // US patent № 0061715 A1. 06.03.2014.

79. Biolek D., Di Ventra M., Pershin Y.V. Reliable SPICE Simulations of Memristors, Memcapacitors and Meminductors // Radioengineering. 2013.V. 22(4) P. 945-968.

80. Zhao W., Cao Y. Predictive Technology Model for Nano-CMOS Design Exploration // ACM Journal on Emerging Technologies in Computing Systems. 2007. V. 3(1). Article 1.

81. Писарев А.Д., Бусыгин А.Н., Бобылев А.Н., Губин А.А., Удовиченко С.Ю. Моделирование и аппаратное тестирование процессов обработки сигналов в комбинированном мемристорно-диодном кроссбаре, являющемся основой биоморфного нейропроцессора // Сборник тезисов 6-ой международной научной конференции «Электронная компонентная база и микроэлектронные модули». Наноиндустрия (Спецвыпуск) 2020. Т. 13. № S4 (99). С. 591-593.

82. Pisarev A.D., Busygin A.N., Bobylev A.N., Gubin A.A., Udovichenko S.Yu. Simulation and experimental evaluation of signal processing routines in a composite memristor-diode crossbar the basis of a biomorphic neuroprocessor // Nanoindustry. 2020. V.13. № S5-3(102). P. 724-732.

83. Писарев А.Д., Бусыгин А.Н., Бобылев А.Н., Губин А.А., Удовиченко С.Ю. Исследование электрофизических свойств комбинированного мемристорно-диодного кроссбара, являющегося основой для аппаратной реализации биоморфного нейропроцессора // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2020. Т.6. №3. С. 93-109.

84. Pisarev A., Busygin A., Bobylev A., Gubin A., Udovichenko S. Fabrication technology and electrophysical properties of a composite memristor-diode crossbar

used as a basis for hardware implementation of a biomorphic neuroprocessor // *Microelectronic Engineering*. 2021. V. 236. Article 111471.

85. Ибрагим А.Х., Бусыгин А.Н., Удовиченко С.Ю. Компьютерная симуляция биоморфного нейропроцессора // Сборник тезисов II Международной научно-технической конференции «Smart energy systems 2021» (SES-2021). С.35,48.

86. Удовиченко С.Ю., Писарев А.Д., Бусыгин А.Н., Ибрагим А.Х., Компьютерная симуляция биоморфного нейропроцессора — носителя искусственного интеллекта // Сборник тезисов международной научной конференции «Применение технологий виртуальной реальности и смежных информационных систем в междисциплинарных задачах FIT-M 2020», 2020, С. 14-19.

87. Udovichenko S.Yu., Pisarev A.D., Busygin A.N., Bobylev A.N. Biomorphous neuroprocessor – prototype of a new generation computer being a carrier of artificial intelligence. Part 2 // *Nanoindustry*. 2021. V. 14. № 1(102). P. 68-79.

88. Бусыгин А.Н., Бобылев А.Н., Губин А.А., Писарев А.Д., Удовиченко С.Ю. Численное моделирование и экспериментальное исследование аппаратной импульсной нейросети с мемристорными синапсами // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2021. Т.7. №2 (26). С. 223-235.

89. Бобылев А.Н., Бусыгин А.Н., Губин А.А., Писарев А.Д., Удовиченко С.Ю. Изготовление и тестирование аппаратной импульсной нейросети с мемристорными синапсами для биоморфного нейропроцессора // Российские нанотехнологии. 2021. Т.16. № 6. С. 793-798

Bobylev A.N., Busygin A.N., Gubin A.A., Pisarev A.D., Udovichenko S.Yu. Manufacture and testing of a pulsed hardware neural network with memristor synapses for a biomorphic neuroprocessor // *Nanobiotechnology Reports*. 2021. V.16. N.6 P. 761-766.

90. Liu J., Mcdaid L.J., Harkin J. et al. Exploring Self-Repair in a Coupled Spiking Astrocyte Neural Network // IEEE Trans. Neural Netw. Learn. Syst. 2018. V. 30. № 3. P. 865.
91. Bobylev A.N., Udovichenko S.Yu. Electrical properties of a $\text{TiN}/\text{Ti}_x\text{Al}_{1-x}\text{O}_y/\text{TiN}$ memristor device manufactured by magnetron sputtering // Russ. Microelectron. 2016. V. 45(6). P. 396.
92. Удовиченко С.Ю., Бусыгин А.Н., Писарев А.Д. Ассоциативное самообучение аппаратной импульсной нейросети, созданной на основе мемристорно-диодного кроссбара // Сборник тезисов XII конференции Нанотехнологического общества России. 2021.

ПРИЛОЖЕНИЯ

РОССИЙСКАЯ ФЕДЕРАЦИЯ

**ПАТЕНТ**

НА ИЗОБРЕТЕНИЕ

№ 2649657

**Запоминающее устройство на основе комплементарной
мемристорно-диодной ячейки**

Патентообладатель: *федеральное государственное автономное
образовательное учреждение высшего образования
"Тюменский государственный университет" (RU)*

Авторы: *Маевский Олег Васильевич (RU), Писарев Александр
Дмитриевич (RU), Бусыгин Александр Николаевич (RU),
Удовиченко Сергей Юрьевич (RU)*

Заявка № 2017109255

Приоритет изобретения 20 марта 2017 г.

Дата государственной регистрации в

Государственном реестре изобретений

Российской Федерации 04 апреля 2018 г.

Срок действия исключительного права

на изобретение истекает 20 марта 2037 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ

(19) **RU** (11) **2 649 657** (13) **C1**(51) МПК
G11C 5/02 (2006.01)ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК
G11C 5/02 (2006.01)

(21)(22) Заявка: 2017109255, 20.03.2017
 (24) Дата начала отсчета срока действия патента:
 20.03.2017
 Дата регистрации:
 04.04.2018
 Приоритет(ы):
 (22) Дата подачи заявки: 20.03.2017
 (45) Опубликовано: 04.04.2018 Бюл. № 10
 Адрес для переписки:
 625003, г. Тюмень, ул. Володарского, 6, ФГАОУ
 ВО "Тюменский государственный университет"

(72) Автор(ы):
 Маевский Олег Васильевич (RU),
 Писарев Александр Дмитриевич (RU),
 Бусыгин Александр Николаевич (RU),
 Удовиченко Сергей Юрьевич (RU)
 (73) Патентообладатель(и):
 федеральное государственное автономное
 образовательное учреждение высшего
 образования "Тюменский государственный
 университет" (RU)
 (56) Список документов, цитированных в отчете
 о поиске: RU 2507611 C1, 20.02.2014. WO
 2016175822 A1, 03.11.2016. US 5646879 A1,
 08.07.1997. US 20160189775 A1, 30.06.2016. WO
 2010087852 A1, 05.08.2010.

(54) Запоминающее устройство на основе комплементарной мемристорно-диодной ячейки

(57) Реферат:

Использование: для построения надежных сверхбольших запоминающих матриц с энергонезависимой памятью, высокой степенью интеграции элементов и малым энергопотреблением. Сущность изобретения заключается в том, что запоминающее устройство на основе комплементарной мемристорно-диодной ячейки, представляющее собой матрицу электрически перепрограммируемых ячеек с параллельным или последовательным доступом к записи и чтению по общим электрическим

шинам, отличается тем, что в каждой ячейке памяти последовательно включены мемристоры и к их общему контакту подключен диод Зенера, так что ячейка имеет подключенные к общим электрическим шинам три вывода, два из которых соединены с контактами мемристоров и еще один - с контактом диода Зенера. Технический результат: обеспечение возможности высокой интеграции при объединении ячеек в сверхбыструю матрицу. 7 ил.

R U 2 6 4 9 6 5 7 C 1

R U 2 6 4 9 6 5 7 C 1

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2697623

**3D запоминающая матрица на основе комплементарной
мемристорно-диодной ячейки**

Патентообладатель: *федеральное государственное автономное
образовательное учреждение высшего образования
"Тюменский государственный университет" (RU)*

Авторы: *Писарев Александр Дмитриевич (RU), Бусыгин
Александр Николаевич (RU), Удовиченко Сергей Юрьевич
(RU), Бобылев Андрей Николаевич (RU), Маевский Олег
Васильевич (RU)*

Заявка № 2017138749

Приоритет изобретения 07 ноября 2017 г.

Дата государственной регистрации в
Государственном реестре изобретений
Российской Федерации 15 августа 2019 г.

Срок действия исключительного права
на изобретение истекает 07 ноября 2037 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(19) RU (11)

2 697 623⁽¹³⁾ C2

(51) МПК
G11C 5/02 (2006.01)
B82B 1/00 (2006.01)

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК
G11C 5/02 (2019.02); B82B 1/00 (2019.02)

(21)(22) Заявка: 2017138749, 07.11.2017
(24) Дата начала отсчета срока действия патента:
07.11.2017
Дата регистрации:
15.08.2019
Приоритет(ы):
(22) Дата подачи заявки: 07.11.2017
(43) Дата публикации заявки: 07.05.2019 Бюл. № 13
(45) Опубликовано: 15.08.2019 Бюл. № 23
Адрес для переписки:
625003, г. Тюмень, ул. Володарского, 6, ФГАОУ
ВО "Тюменский государственный университет"

(72) Автор(ы):
Писарев Александр Дмитриевич (RU),
Бусыгин Александр Николаевич (RU),
Удовиченко Сергей Юрьевич (RU),
Бобылев Андрей Николаевич (RU),
Маевский Олег Васильевич (RU)
(73) Патентообладатель(и):
федеральное государственное автономное
образовательное учреждение высшего
образования "Тюменский государственный
университет" (RU)
(56) Список документов, цитированных в отчете
о поиске: WO 2017/131628 A1, 03.08.2017. WO
2017/039611 A1, 09.03.2017. T. POTTEIGER et
al. "A one Zener diode, one memristor crossbar
architecture for a write-time-based PUF", опубл.
01.10.2015 на 4 страницах [найдено 21.02.2019],
найдено в Интернет по адресу URL:
<https://ieeexplore.ieee.org/document/7282123>. Г.
КРИВУЛЯ и др. "Особенности применения
(см. прод.)

(54) 3D запоминающая матрица на основе комплементарной мемристорно-диодной ячейки

(57) Реферат:

Изобретение относится к области микро- и наноэлектроники. Технический результат заключается в обеспечении высокой степени интеграции элементов, малого энергопотребления и высокого быстродействия устройства. 3D запоминающая матрица на основе комплементарной мемристорно-диодной ячейки, представляющая собой электронное интегральное устройство с энергонезависимой памятью, причем электронное интегральное устройство с энергонезависимой памятью представляет собой 3D структуру, образованную из нанесенных друг на друга комбинированных кроссбаров,

состоящих из запоминающих ячеек, содержащих два мемристора с общим электродом, соединенным с одним из контактов диода Зенера, и образующихся в пересечении двух параллельных проводников с одной стороны кроссбара, объединяющих соответствующие электроды мемристорных ячеек в столбцы, и одного ортогонального к ним проводника с другой стороны, объединяющего контакты диодов Зенера соседних ячеек в строки, причем соседние кроссбары ориентированы зеркально по отношению друг к другу и имеют общие шины строк или столбцов. 5 ил.

RU 2 697 623 C 2

RU 2 697 623 C 2